

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>5</sup>:</b> <b>C12N 9/22, 15/55, 15/74</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 94/18313</b> <b>(43) International Publication Date:</b> 18 August 1994 (18.08.94)
<b>(21) International Application Number:</b> PCT/US94/01201 <b>(22) International Filing Date:</b> 10 February 1994 (10.02.94)  <b>(30) Priority Data:</b> 08/017,493 12 February 1993 (12.02.93) US  <b>(71) Applicant:</b> THE JOHNS-HOPKINS UNIVERSITY [US/US]; 34th & Charles Streets, Baltimore, MD 21218 (US).  <b>(72) Inventor:</b> CHANDRASEGARAN, Srinivasan; 4 East 32nd Street, #206, Baltimore, MD 21218 (US).  <b>(74) Agents:</b> KOKULIS, Paul, N. et al.; Cushman, Darby & Cushman, 1100 New York Avenue, N.W., Washington, DC 20005 (US).		<b>(81) Designated States:</b> AU, CA, JP, NZ, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>With international search report.</i> <i>With amended claims.</i>
<b>(54) Title:</b> FUNCTIONAL DOMAINS IN <i>FLAVOBACTERIUM OKEANOKOITES</i> (FOKI) RESTRICTION ENDONUCLEASE		
<b>(57) Abstract</b>  The present inventors have identified the recognition and cleavage domains of the <i>FokI</i> restriction endonuclease. Accordingly, the present invention relates to the DNA segments encoding the recognition and cleavage domains of the <i>FokI</i> restriction endonuclease, respectively. The 41 kDa N-terminal fragment constitutes the <i>FokI</i> recognition domain while the 25 kDa C-terminal fragment constitutes the <i>FokI</i> cleavage nuclease domain. The present invention also relates to hybrid restriction enzymes comprising the nuclease domain of the <i>FokI</i> restriction endonuclease linked to a recognition domain of another enzyme. Additionally, the present invention relates to the construction of two insertion mutants of <i>FokI</i> endonuclease.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

**FUNCTIONAL DOMAINS IN FLAVOBACTERIUM OKEANOKOITES  
(FOKI) RESTRICTION ENDONUCLEASE**

**BACKGROUND OF THE INVENTION**

**1. Field of the Invention:**

5           The present invention relates to the FokI restriction endonuclease system. In particular, the present invention relates to DNA segments encoding the separate functional domains of this restriction endonuclease system.

10           The present invention also relates to the construction of two insertion mutants of FokI endonuclease.

**2. Background Information:**

15           Type II endonucleases and modification methylases are bacterial enzymes that recognize specific sequences in duplex DNA. The endonuclease cleaves the DNA while the methylases methylates adenine or cytosine residues so as to protect the host-genome against cleavage [Type II restriction and modification enzymes. In Nucleases (Eds. Modrich  
20           and Roberts) Cold Spring Harbor Laboratory, New York, pp. 109-154, 1982]. These restriction-modification (R-M) systems function to protect cells from infection by phage and plasmid molecules that  
25           would otherwise destroy them.

As many as 2500 restriction enzymes with over 200 specificities have been detected and purified (Wilson and Murray, Annu. Rev. Genet. 25:585-627, 1991). The recognition sites of most of these enzymes are 4-6 base pairs long. The small size of the recognition sites is beneficial as the phage genomes are usually small and these small recognition sites occur more frequently in the phage.

Eighty different R-M systems belonging to the Type IIS class with over 35 specificities have been identified. This class is unique in that the cleavage site of the enzyme is separate from the recognition sequence. Usually the distance between the recognition site and the cleavage site is quite precise (Szybalski et al., Gene, 100:13-26, 1991). Among all these enzymes, the *FokI* restriction endonuclease is the most well characterized member of the Type IIS class. The *FokI* endonuclease (*RFokI*) recognizes asymmetric pentanucleotides in double-stranded DNA, 5' GGATG-3' (SEQ ID NO: 1) in one strand and 3'-CCTAC-5' (SEQ ID NO: 2) in the other, and introduces staggered cleavages at sites away from the recognition site (Sugisaki et al., Gene 16:73-78; 1981). In contrast, the *FokI* methylase (*MFokI*) modifies DNA thereby rendering the DNA resistant to digestion by *FokI* endonuclease. The *FokI* restriction and modification genes have been cloned and their nucleotide sequences deduced (Kita et al., J. of Biol. Chem., 264:575-5756, 1989). Nevertheless, the domain structure of the *FokI* restriction endonuclease remains unknown, although a three domain structure has been suggested

(Wilson and Murray, Annu. Rev. Genet. 25:585-627, 1991).

#### SUMMARY OF THE INVENTION

Accordingly, it is an object of the present invention to provide isolated domains of Type IIS restriction endonuclease.

It is another object of the present invention to provide hybrid restriction enzymes which are useful for mapping and sequencing.

An additional object of the present invention is to provide two insertion mutants of *FOXI* which have an increased distance of cleavage from the recognition site as compared to the wild-type enzyme. The polymerase chain reaction (PCR) is utilized to construct the two mutants.

Various other objects and advantages of the present invention will become obvious from the drawings and the following description of the invention.

In one embodiment, the present invention relates to a DNA segment encoding the recognition domain of a Type IIS endonuclease which contains the sequence-specific recognition activity of the Type IIS endonuclease or a DNA segment encoding the catalytic domain of a Type IIS endonuclease which contains the cleavage activity of the Type IIS endonuclease.

In another embodiment, the present invention relates to an isolated protein consisting essentially of the N-terminus or recognition domain of the *FokI* restriction endonuclease which protein has the sequence-specific recognition activity of the endonuclease or an isolated protein consisting

essentially of the C-terminus or catalytic domain of the *FokI* restriction endonuclease which protein has the nuclease activity of the endonuclease.

In a further embodiment, the present invention relates to a DNA construct comprising a first DNA segment encoding the catalytic domain of a Type IIS endonuclease which contains the cleavage activity of the Type IIS endonuclease; a second DNA segment encoding a sequence-specific recognition domain other than the recognition domain of the Type IIS endonuclease; and a vector. In the construct, the first DNA segment and the second DNA segment are operably linked to the vector to result in the production of a hybrid restriction enzyme.

In another embodiment, the present invention relates to a hybrid restriction enzyme comprising the catalytic domain of a Type IIS endonuclease which contains the cleavage activity of the Type IIS endonuclease linked to a recognition domain of an enzyme or a protein other than the Type IIS endonuclease from which the cleavage domain is obtained.

In a further embodiment, the present invention relates to a DNA construct comprising a first DNA segment encoding the catalytic domain of a Type IIS endonuclease which contains the cleavage activity of the Type IIS endonuclease; a second DNA segment encoding a sequence-specific recognition domain other than the recognition domain of the Type IIS endonuclease; a third DNA segment comprising one or more codons, wherein the third DNA segment is inserted between the first DNA segment and the

second DNA segment; and a vector. Preferably, the third segment contains four or seven codons.

In another embodiment, the present invention relates to a procaryotic cell comprising a first DNA segment encoding the catalytic domain of a Type IIS endonuclease which contains the cleavage activity of the Type IIS endonuclease; a second DNA segment encoding a sequence-specific recognition domain other than the recognition domain of the Type IIS endonuclease; a third DNA segment comprising one or more codons, wherein the third DNA segment is inserted between the first DNA segment and the second DNA segment; and a vector. The first DNA segment and the second DNA segment are operably linked to the vector so that a single protein is produced.

#### BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1 shows sequences of the 5' and 3' primers used to introduce new translation signals into *fokIM* and *fokIR* genes during PCR amplification. (SEQ ID NOs: 3-9). SD represents Shine-Dalgarno consensus RBS for *Escherichia coli* (*E. coli*) and 7-bp spacer separates the RBS from the ATG start codon. The *fokIM* primers are flanked by *NcoI* sites. The *fokIR* primers are flanked by *BamHI* sites. Start and stop codons are shown in bold letters. The 18-bp complement sequence is complementary to the sequence immediately following the stop codon of *MfokI* gene.

FIGURE 2 shows the structure of plasmids pACYCM*fokIM*, pRRSR*fokIR* and pCB*fokIR*. The PCR-modified *fokIM* gene was inserted at the *NcoI* site of pACYC184 to form pACYC*fokIM*. The PCR-generated

*fokIR* gene was inserted at the *Bam*HI sites of pRRS and pCB to form pRRS*fokIR* and pCB*fokIR*, respectively. pRRS possesses a *lac* UV5 promoter and pCB contains a strong *tac* promoter. In addition, these vectors contain the positive retroregulator sequence downstream of the inserted *fokIR* gene.

FIGURE 3 shows SDS (0.1%) - polyacrylamide (12%) gel electrophoretic profiles at each step in the purification of *FokI* endonuclease. Lanes: 1, protein standards; 2, crude extract from uninduced cells; 3, crude extract from cells induced with 1 mM IPTG; 4, phosphocellulose pool; 5, 50-70% (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> fractionation pool; and 6, DEAE pool.

FIGURE 4 shows SDS (0.1%) - polyacrylamide (12%) gel electrophoretic profiles of tryptic fragments at various time points of trypsin digestion of *FokI* endonuclease in presence of the oligonucleotide DNA substrate, d-5'-CCTCTGGATGCTCTC-3' (SEQ ID NO: 10): 5'-GAGAGCATCCAGAGG-3' (SEQ ID NO: 11). Lanes: 1, protein standards; 2, *FokI* endonuclease; 3, 2.5 min; 4, 5 min; 5, 10 min; 6, 20 min; 7, 40 min; 8, 80 min; 9, 160 min of trypsin digestion respectively. Lanes 10-13: HPLC purified tryptic fragments. Lanes: 10, 41 kDa fragment; 11, 30 kDa fragment; 12, 11 kDa fragment; and 13, 25 kDa fragment.

FIGURE 5 shows the identification of DNA binding tryptic fragments of *FokI* endonuclease using an oligo dT-cellulose column. Lanes: 1, protein standards, 2, *FokI* endonuclease; 3, 10 min trypsin digestion mixture of *FokI* - oligo complex; 4, tryptic fragments that bound to the oligo dT-cellulose column; 5, 160 min trypsin digestion



mixture of *FokI* - oligo complex; 6, tryptic fragments that bound to the oligo dT-cellulose column.

FIGURE 6 shows an analysis of the cleavage properties of the tryptic fragments of *FokI* endonuclease.

(A) The cleavage properties of the tryptic fragments were analyzed by agarose gel electrophoresis. 1  $\mu$ g of pTZ19R in 10mM Tris.HCl (pH 8), 50mM NaCl, 1mM DTT, and 10mM MgCl<sub>2</sub> was digested with 2  $\mu$ l of the solution containing the fragments (tryptic digests, breakthrough and eluate respectively) at 37°C for 1 hr in a reaction volume of 10  $\mu$ l. Lanes 4 to 6 correspond to trypsin digestion of *FokI* - oligo complex in absence of MgCl<sub>2</sub>. Lanes 7 to 9 correspond to trypsin digestion of *FokI* - oligo complex in presence of 10 mM MgCl<sub>2</sub>. Lanes: 1, 1 kb ladder; 2, pTZ19R; 3, pTZ19R digested with *FokI* endonuclease; 4 and 6, reaction mixture of the tryptic digests of *FokI* - oligo complex; 5 and 7, 25 kDa C-terminal fragment in the breakthrough volume; 6 and 9, tryptic fragments of *FokI* that bound to the DEAE column. The intense bands at bottom of the gel correspond to excess oligonucleotides.

(B) SDS (0.1%) - polyacrylamide (12%) gel electrophoretic profiles of fragments from the DEAE column. Lanes 3 to 5 correspond to trypsin digestion of *FokI* - oligo complex in absence of MgCl<sub>2</sub>. Lanes 6 to 8 correspond to trypsin digestion of *FokI* - oligo complex in presence of 10 mM MgCl<sub>2</sub>. Lanes: 1, protein standards; 2, *FokI* endonuclease; 3 and 6, reaction mixture of the tryptic digests of

*FokI* - oligo complex; 4 and 7, 25 kDa C-terminal fragment in the breakthrough volume; 5 and 8, tryptic fragments of *FokI* that bound to the DEAE column.

5                   FIGURE 7 shows an analysis of sequence -  
specific binding of DNA by 41 kDa N-terminal  
fragment using gel mobility shift assays. For the  
exchange reaction, the complex (10  $\mu$ l) was incubated  
with 1  $\mu$ l of  $^{32}$ P-labeled specific (or non-specific)  
10 oligonucleotide duplex in a volume of 20  $\mu$ l  
containing 10 mM Tris.HCl, 50 mM NaCl and 10 mM MgCl<sub>2</sub>  
at 37°C for various times. 1  $\mu$ l of the 5'- $^{32}$ P-  
labeled specific probe [d-5'-CCTCTGGATGCTCTC-3' (SEQ  
ID NO: 10): 5'-GAGAGCATCCAGAGG-3' (SEQ ID NO: 11)]  
15 contained 12 picomoles of the duplex and  $\sim 50 \times 10^3$   
cpm. 1  $\mu$ l of the 5'- $^{32}$ P-labeled non-specific probe  
[5'-TAATTGATTCTTAA-3' (SEQ ID NO: 12): 5'-  
ATTAAGAATCAATT-3' (SEQ ID NO: 13)] contained 12  
picomoles of the duplex and  $\sim 25 \times 10^3$  cpm. (A)  
20 Lanes: 1, specific oligonucleotide duplex; 2, 41  
kDa N-terminal fragment-oligo complex; 3 and 4,  
specific probe incubated with the complex for 30 and  
120 min respectively. (B) Lanes: 1, non-specific  
oligonucleotide duplex; 2, 41 kDa N-terminal  
25 fragment-oligo complex; 3 and 4 non-specific probe  
incubated with the complex for 30 and 120 min  
respectively.

FIGURE 8 shows SDS (0.1%) polyacrylamide  
(12%) gel electrophoretic profiles of tryptic  
30 fragments at various time points of trypsin  
digestion of *FokI* endonuclease. The enzyme (200  $\mu$ g)  
in a final volume of 200  $\mu$ l containing 10 mM  
Tris.HCl, 50 mM NaCl and 10mM MgCl<sub>2</sub> was digested with

trypsin at RT. The trypsin to *FokI* ratio was 1:50 by weight. Aliquots (28  $\mu$ l) from the reaction mixture removed at different time intervals and quenched with excess antipain. Lanes: 1, protein standards; 2, *FokI* endonuclease; 3, 2.5 min; 4, 5.0 min; 5, 10 min; 6, 20 min; 7, 40 min; 8, 80 min; and 9, 160 min of trypsin digestion respectively.

FIGURE 9 shows the tryptic map of *FokI* endonuclease (A) *FokI* endonuclease fragmentation pattern in absence of the oligonucleotide substrate. (B) *FokI* endonuclease fragmentation pattern in presence of the oligonucleotide substrate.

FIGURE 10 shows the predicted secondary structure of *FokI* based on its primary sequencing using the PREDICT program. (See SEQ ID NO:31) The trypsin cleavage site of *FokI* in the presence of DNA substrates is indicated by the arrow. The KSELEEKKSEL segment is highlighted. The symbols are as follows: h, helix; s, sheet; and ., random coil.

FIGURE 11 shows the sequences of the 5' and 3' oligonucleotide primers used to construct the insertion mutants of *FokI* (see SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38 and SEQ ID NO:39, respectively). The four and seven codon inserts are shown in bold letters. The amino acid sequence is indicated over the nucleotide sequence. The same 3' primer was used in the PCR amplification of both insertion mutants.

FIGURE 12 shows the SDS/PAGE profiles of the mutant enzymes purified to homogeneity. Lanes: 1, protein standards; 2, *FokI*; 3, mutant *FokI* with

4-codon insertion; and 4, mutant *FokI* with 7-codon insertion.

FIGURE 13 shows an analysis of the DNA sequence specificity of the mutant enzymes. The DNA substrates were digested in 10 mM Tris HCl, pH 8.0/50 mM NaCl/1 mM DTT/10mM MgCl<sub>2</sub> at 37°C for 2 hrs.

(A) Cleavage pattern of pTZ19R DNA substrate analyzed by 1% agarose gel electrophoresis. 2μg of pTZ19R DNA was used in each reaction. Lanes: 1, 1-kilobase (kb) ladder; 2, pTZ19R; 3, pTZ19R digested with *FokI*; pTZ19R digested with mutant *FokI* with 4-codon insertion; and 5, pTZ19R digested with mutant *FokI* with 7-codon insertion.

(B) Cleavage pattern of 256 bp DNA substrate containing a single *FokI* site analyzed by 1.5% agarose gel electrophoresis. 1μg of radiolabeled substrates (<sup>32</sup>P-labeled on individual strands) was digested as described above. The agarose gel was stained with ethidium bromide and visualized under UV light. Lanes 2 to 6 correspond to the <sup>32</sup>P-labeled substrate in which the 5'-CATCC-3' strand is <sup>32</sup>-P labeled. Lanes 7 to 11 correspond to the substrate in which the 5'-GGATG-3' strand is <sup>32</sup>p-labeled. Lanes: 1, 1kb ladder; 2 and 7, <sup>32</sup>P-labeled 250 bp DNA substrates; 3 and 8, <sup>32</sup>-P labeled substrates cleaved with *FokI*; 4 and 9, purified the laboratory wild-type *FokI*; 5 and 10, mutant *FokI* with 4-codon insertion; 6 and 11, mutant *FokI* with 7-codon insertion.

(C) Autoradiograph of the agarose gel from above. Lanes: 2 to 11, same as in B.

FIGURE 14 shows an analysis of the distance of cleavage from the recognition site by *FokI* and the mutant enzymes. The unphosphorylated oligonucleotides were used for dideoxy DNA sequencing with pTZ19R as the template. The sequencing products (G, A, T, C) were electrophoresed on a 6% acrylamide gel containing 7M urea, and the gel dried. The products were then exposed to an x-ray film for 2 hrs. Cleavage products from the 100 bp and the 256 bp DNA substrates are shown in A and B, respectively. I corresponds to substrates containing <sup>32</sup>P-label on the 5'-GGATG-3' strand, and II corresponds to substrates containing <sup>32</sup>P-label on the 5'-CATCC-3' strand. Lanes: 1, *FokI*; 2, *FokI* ; 3, mutant *FokI* with 4-codon insertion; and 4, mutant *FokI* with 7-codon insertion.

FIGURE 15 shows a map of the cleavage site(s) of *FokI* and the mutant enzymes based on the 100 bp DNA substrate containing a single *FokI* site: (A) wild-type *FokI*; (B) mutant *FokI* with 4-codon insertion; and (C) mutant *FokI* with 7-codon insertion (see SEQ ID NO:40). The sites of cleavage are indicated by the arrows. Major cleavage sites are shown by larger arrows.

#### DETAILED DESCRIPTION OF THE INVENTION

The present invention is based on the identification and characterization of the functional domains of the *FokI* restriction endonuclease. In the experiments resulting in the present invention, it was discovered that the *FokI* restriction endonuclease is a two domain system, one domain of which possesses the sequence-specific

recognition activity while the other domain contains the nuclease cleavage activity.

The *FokI* restriction endonuclease recognizes the non-palindromic pentanucleotide 5'-GGATG-3' (SEQ ID NO: 1):5'-CATCC-3' (SEQ ID NO: 2) in duplex DNA and cleaves 9/13 nucleotides downstream from the recognition site. Since 10 base pairs are required for one turn of the DNA helix, the present inventors hypothesized that the enzyme would interact with one face of the DNA by binding at one point and cleave at another point on the next turn of the helix. This suggested the presence of two separate protein domains, one for sequence-specific recognition of DNA and one for endonuclease activity. The hypothesized two domain structure was shown to be the correct structure of the *FokI* endonuclease system by studies that resulted in the present invention.

Accordingly, in one embodiment, the present invention relates to a DNA segment which encodes the N-terminus of the *FokI* restriction endonuclease (preferably, about the N-terminal 2/3's of the protein). This DNA segment encodes a protein which has the sequence-specific recognition activity of the endonuclease, that is, the encoded protein recognizes the non-palindromic pentanucleotide d-5'-GGATG-3' (SEQ ID NO: 1):5'-CATCC-3' (SEQ ID NO: 2) in duplex DNA. Preferably, the DNA segment of the present invention encodes amino acids 1-382 of the *FokI* endonuclease.

In a further embodiment, the present invention relates to a DNA segment which encodes the C-terminus of the *FokI* restriction endonuclease.

The protein encoded by this DNA segment of the present invention has the nuclease cleavage activity of the *FokI* restriction endonuclease. Preferably, the DNA segment of the present invention encodes amino acids 383-578 of the *FokI* endonuclease. DNA segments of the present invention can be readily isolated from a biological samples using methods known in the art, for example, gel electrophoresis, affinity chromatography, polymerase chain reaction (PCR) or a combination thereof. Further, the DNA segments of the present invention can be chemically synthesized using standard methods in the art.

The present invention also relates to the proteins encoded by the DNA segments of the present invention. Thus, in another embodiment, the present invention relates to a protein consisting essentially of the N-terminus of the *FokI* endonuclease which retains the sequence-specific recognition activity of the enzyme. This protein of the present invention has a molecular weight of about 41 kilodaltons as determined by SDS polyacrylamide gel electrophoresis in the presence of 2-mercaptoethanol.

In a further embodiment, the present invention relates to a protein consisting essentially of the C-terminus of the *FokI* restriction endonuclease (preferably, the C-terminal 1/3 of the protein). The molecular weight of this protein is about 25 kilodaltons as determined by SDS polyacrylamide gel electrophoresis in the presence of 2-mercaptoethanol.

The proteins of the present invention can be isolated or purified from a biological sample

using methods known in the art. For example, the proteins can be obtained by isolating and cleaving the *FokI* restriction endonuclease. Alternatively, the proteins of the present invention can be

5 chemically synthesized or produced using recombinant DNA technology and purified.

The DNA segments of the present invention can be used to generate 'hybrid' restriction enzymes by linking other DNA binding protein domains with

10 the nuclease domain of *FokI*. This can be achieved chemically as well as by recombinant DNA technology. Such chimeric enzymes are useful for physical mapping and sequencing of genomes of various species, such as, humans, mice and plants. For

15 example, such enzymes would be suitable for use in mapping the human genome.

Such chimeric enzymes are also valuable research tools in recombinant DNA technology and molecular biology. Currently only 4-6 base pair

20 cutters and a few 8 base pair cutters are available commercially. (There are about 10 endonucleases which cut >6 base pairs that are available commercially.) By linking other DNA binding proteins to the nuclease domain of *FokI* new enzymes

25 can be generated that recognize more than 6 base pairs in DNA.

Accordingly, in a further embodiment, the present invention relates to a DNA construct and the hybrid restriction enzyme encoded therein. The

30 DNA construct of the present invention comprises a first DNA segment encoding the nuclease domain of the *FokI* restriction endonuclease, a second DNA segment encoding a sequence-specific recognition



domain and a vector. The first DNA segment and the second DNA segment are operably linked to the vector so that expression of the segments can be effected thereby yielding a chimeric restriction enzyme. The construct can comprise regulatory elements such as promoters (for example, T7, *tac*, *trp* and *lac UV5* promoters), transcriptional terminators or retroregulators (for example, stem loops). Host cells (procaryotes such as *E. coli*) can be transformed with the DNA constructs of the present invention and used for the production of chimeric restriction enzymes.

The hybrid enzymes of the present invention comprise the nuclease domain of *FokI* linked to a recognition domain of another enzyme or DNA binding protein (such as, naturally occurring DNA binding proteins that recognize 6 base pairs). Suitable recognition domains include, but are not limited to, the recognition domains of zinc finger motifs; homeo domain motifs; other DNA binding protein domains of *lambda* repressor, *lac* repressor, *cro*, *gal4*; DNA binding protein domains of oncogenes such as *myc*, *jun*; and other naturally occurring sequence-specific DNA binding proteins that recognize >6 base pairs.

The hybrid restriction enzymes of the present invention can be produced by those skilled in the art using known methodology. For example, the enzymes can be chemically synthesized or produced using recombinant DNA technology well known in the art. The hybrid enzymes of the present invention can be produced by culturing host cells (such as, HB101, RR1, RB791 and MM294) containing

the DNA construct of the present invention and isolating the protein. Further, the hybrid enzymes can be chemically synthesized, for example, by linking the nuclease domain of the *FokI* to the recognition domain using common linkage methods known in the art, for example, using protein cross-linking agents such as EDC/NHS, DSP, etc.

While the *FokI* restriction endonuclease was the enzyme studied in the following experiments, it is expected that other Type IIS endonucleases (such as, those listed in Table 2) will function using a similar two domain structure which one skilled in the art could readily determine based on the present invention.

Recently, *StsI*, a heteroschizomer of *FokI* has been isolated from *Streptococcus sanguis* (Kita et al., Nucleic Acids Research 20 (3) 618, 1992). *StsI* recognizes the same nonpalindromic pentadeoxyribonucleotide 5'-GGATG-3':5'-CATCC-3' as *FokI* but cleaves 10/14 nucleotides downstream of the recognition site. The *StsI* RM system has been cloned and sequenced (Kita et al., Nucleic Acids Research 20 (16) 4167-72, 1992). Considerable amino acid sequence homology (~30%) has been detected between the endonucleases, *FokI* and *StsI*.

Another embodiment of the invention relates to the construction of two insertion mutants of *FokI* endonuclease using the polymerase chain reaction (PCR). In particular, this embodiment includes a DNA construct comprising a first DNA segment encoding the catalytic domain of a Type IIS endonuclease which contains the cleavage activity of the Type IIS endonuclease, a second DNA segment

encoding a sequence-specific recognition domain other than the recognition domain of the Type IIS endonuclease, and a third DNA segment comprising one or more codons. The third DNA segment is inserted  
5 between the first DNA segment and the second DNA segment. The construct also includes a vector. The Type IIS endonuclease is *FokI* restriction endonuclease.

Suitable recognition domains include, but  
10 are not limited to, zinc finger motifs, homeo domain motifs, DNA binding domains of repressors, DNA binding domains of oncogenes and naturally occurring sequence-specific DNA binding proteins that recognize >6 base pairs.

As noted above, the recognition domain of  
15 *FokI* restriction endonuclease is at the amino terminus of *FokI* endonuclease, whereas the cleavage domain is probably at the carboxyl terminal third of the molecule. It is likely that the domains are  
20 connected by a linker region, which defines the spacing between the recognition and the cleavage sites of the DNA substrate. This linker region of *FokI* is susceptible to cleavage by trypsin in the presence of a DNA substrate yielding a 41-kDa amino-  
25 terminal fragment (The DNA binding domain) and a 25-kDa carboxyl-terminal fragment (the cleavage domain). Secondary structure prediction of *FokI* endonuclease based on its primary amino acid sequence supports this hypothesis (see Figure 10).  
30 The predicted structure reveals a long stretch of alpha helix region at the junction of the recognition and cleavage domains. This helix probably constitutes the linker which connects the

two domains of the enzyme. Thus, it was thought that the cleavage distance of *FokI* from the recognition site could be altered by changing the length of this spacer (the alpha helix). Since 3.6  
5 amino acids are required to form one turn of the alpha helix, insertion of either four codons or seven codons in this region would extend the pre-existing helix in the native enzyme by one or two turns, respectively. Close examination of the amino  
10 acid sequence of this helix region revealed the presence of two KSEL repeats separated by amino acids EEK (Figure 10) (see SEQ ID NO:21). The segments KSEL (4 codons) (see SEQ ID NO:22) and KSELEEK (7 codons) (see SEQ ID NO:23) appeared to be  
15 good choices for insertion within this helix in order to extend it by one and two turns, respectively. (See Examples X and XI.) Thus, genetic engineering was utilized in order to create mutant enzymes.

20 In particular, the mutants are obtained by inserting one or more, and preferably four or seven, codons between the recognition and cleavage domains of *FokI*. More specifically, the four or seven codons are inserted at nucleotide 1152 of the gene  
25 encoding the endonuclease. The mutants have the same DNA sequence specificity as the wild-type enzyme. However, they cleave one nucleotide further away from the recognition site on both strands of the DNA substrates as compared to the wild-type  
30 enzyme.

Analysis of the cut sites of *FokI* and the mutants, based on the cleavage of the 100 bp fragment, is summarized in Figure 15. Insertion of

four (or seven) codons between the recognition and cleavage domains of *FokI* is accompanied by an increase in the distance of cleavage from the recognition site. This information further supports the presence of two separate protein domains within the *FokI* endonuclease: one for the sequence specific recognition and the other for the endonuclease activity. The two domains are connected by a linker region which defines the spacing between the recognition and the cleavage sites of the DNA substrate. The modular structure of the enzyme suggests it may be feasible to construct chimeric endonucleases of different sequence specificity by linking other DNA-binding proteins to the cleavage domain of the *FokI* endonuclease.

In view of the above-information, another embodiment of the invention includes a procaryotic cell comprising a first DNA segment encoding the catalytic domain of a Type IIS endonuclease which contains the cleavage activity of the Type IIS endonuclease, a second DNA segment encoding a sequence-specific recognition domain other than the recognition domain of the Type IIS endonuclease, and a third DNA segment comprising one or more codons. The third DNA segment is inserted between the first DNA segment and the second DNA segment. The cell also includes a vector. Additionally, it should be noted that the first DNA segment and the second DNA segment are operably linked to the vector so that a single protein is produced. The third segment may consist essentially of four or seven codons.

The present invention also includes the protein produced by the procaryotic cell referred to directly above. In particular, the isolated protein consists essentially of the recognition domain of the FokI restriction endonuclease, the catalytic domain of the FokI restriction endonuclease, and amino acids encoded by the codons present in the third DNA segment.

The following non-limiting Examples are provided to describe the present invention in greater detail.

#### EXAMPLES

The following materials and methods were utilized in the isolation and characterization of the FokI restriction endonuclease functional domains as exemplified hereinbelow.

##### Bacterial strains and plasmids

Recombinant plasmids were transformed into *E.coli* RB791 *i<sup>q</sup>* cells which carry the *lac i<sup>q</sup>* allele on the chromosome (Brent and Ptashne, PNAS USA, 78:4204-4208, 1981) or *E.coli* RR1 cells. Plasmid pACYCfokIM is a derivative of pACYC184 carrying the PCR-generated *fokIM* gene inserted into *NcoI* site. The plasmid expresses the FokI methylase constitutively and was present in RB791 cells (or RR1 cells) whenever the *fokIR* gene was introduced on a separate compatible plasmid. The FokI methylase modifies FokI sites and provides protection against chromosomal cleavage. The construction of vectors pRRS and pCB are described elsewhere (Skoglund et al., Gene, 88:1-5, 1990).

Enzymes, biochemicals and oligos

Oligo primers for PCR were synthesized with an Applied Biosystem DNA synthesizer using cyanoethyl phosphoramidite chemistry and purified by reversed phase HPLC. Restriction enzymes were purchased from New England Biolabs. The DNA ligase IPTG were from Boehringer-Mannheim. PCR reagents were purchased as a Gene Amp Kit from Perkin-Elmer. Plasmid purification kit was from QIAGEN.

Restriction enzyme assays

Cells from a 5-ml sample of culture medium were harvested by centrifugation, resuspended in 0.5 ml sonication buffer [50 mM Tris.HCl (pH 8), 14mM 2-mercaptoethanol], and disrupted by sonication (3 x 5 seconds each) on ice. The cellular debris was centrifuged and the crude extract used in the enzyme assay. Reaction mixtures (10  $\mu$ l) contained 10mM Tris.HCl (pH 8), 10 mM MgCl<sub>2</sub>, 7 mM 2-mercaptoethanol, 50  $\mu$ g of BSA, 1  $\mu$ g of plasmid pTZ19R (U.S. biochemicals) and 1 $\mu$ l of crude enzyme. Incubation was at 37°C for 15 min. tRNA (10  $\mu$ g) was added to the reaction mixtures when necessary to inhibit non-specific nucleases. After digestion, 1  $\mu$ l of dye solution (100 mM EDTA, 0.1% bromophenol blue, 0.1% xylene cyanol, 50% glycerol) was added, and the samples were electrophoresed on a 1% agarose gel. Bands were stained with 0.5  $\mu$ g ethidium bromide/ml and visualized with 310-nm ultraviolet light.

SDS/PAGE

Proteins were prepared in sample buffer and electrophoresed in SDS (0.1%)- polyacrylamide (12%) gels as described by Laemmli (Laemmli, Nature,

222:680-685, 1970). Proteins were stained with coomassie blue.

#### Example I

##### Cloning of FokI RM system

5           The FokI system was cloned by selecting for the modification phenotype. *Flavobacterium okeanokoites* strain DNA was isolated by the method described by Caserta et al. (Caserta et al., J. Biol. Chem., 262:4770-4777, 1987). Several  
10 *Flavobacterium okeanokoites* genome libraries were constructed in plasmids pBR322 and pUC13 using the cloning enzymes *PstI*, *BamHI* and *BglII*. Plasmid library DNA (10 µg) was digested with 100 units of  
15 *FokI* endonuclease to select for plasmids expressing *fokIM*+ phenotype.

          Surviving plasmids were transformed into RR1 cells and transformants were selected on plates containing appropriate antibiotic. After two rounds of biochemical enrichment, several plasmids  
20 expressing the *fokIM*+ phenotype from these libraries were identified. Plasmids from these clones were totally resistant to digestion by *FokI*.

          Among eight transformants that were analyzed from the *F. okeanokoites* pBR322 *PstI*  
25 library, two appeared to carry the *fokIM* gene and plasmids from these contained a 5.5 kb *PstI* fragment. Among eight transformants that were picked from *F. okeanokoites* pBR322 *BamHI* library, two appeared to carry the *fokIM* gene and their  
30 plasmids contained ~ 18 kb *BamHI* fragment. Among eight transformants that were analyzed from the *F. okeanokoites* genome *BglII* library in pUC13, six appeared to carry the *fokIM* gene. Three of these



clones had a 8 kb *Bgl*III insert while the rest contained a 16 kb *Bgl*III fragment.

Plating efficiency of phage  $\lambda$  on these clones suggested that they also carried the *fokIR* gene. The clones with the 8-kb *Bgl*III insert appeared to be most resistant to phage infection. Furthermore, the *Fok*I endonuclease activity was detected in the crude extract of this clone after partial purification on a phosphocellulose column. The plasmid, pUC*fokIRM* from this clone was chosen for further characterization.

The 5.5 kb *Pst*I fragment was transferred to M13 phages and the nucleotide sequences of parts of this insert determined using Sanger's sequencing method (Sanger et al., PNAS USA, 74:5463-5467, 1977). The complete nucleotide sequence of the *Fok*I RM system has been published by other laboratories (Looney et al., Gene, 80:193-208, 1989; Kita et al., Nucleic Acid Res., 17:8741-8753, 1989; Kita et al., J. Biol. Chem. 264:5751-5756, 1989).

#### Example II

##### Construction of an efficient overproducer clone of *Fok*I endonuclease using polymerase chain reaction.

The PCR technique was used to alter transcriptional and translational signals surrounding the *fokIR* gene so as to achieve overexpression in *E. coli* (Skoglund et al., Gene, 88:1-5, 1990). The ribosome-binding site preceding the *fokIR* and *fokIM* genes were altered to match the consensus *E. coli* signal.

In the PCR reaction, plasmid pUC*fokIRM* DNA linearized with *Bam*HI was used as the template. PCR reactions (100  $\mu$ l) contained 0.25 nmol of each

primer, 50  $\mu$ M of each dNTP, 10 mM Tris.HCl (pH 8.3 at 25°C), 50 mM KCl, 1.5 mM MgCl<sub>2</sub> 0.01% (W/V) gelatin, 1 ng of template DNA, 5 units of Taq DNA polymerase. The oligo primes used for the amplification of the *fokIR* and *fokIM* genes are shown in Figure 1. Reaction mixtures (ran in quadruplicate) were overlayed with mineral oil and reactions were carried out using Perkin-Elmer-Cetus Thermal Cyclor.

Initial template denaturation was programmed for 2 min. Thereafter, the cycle profile was programmed as follows: 2 min at 37°C (annealing), 5 min at 72°C (extension), and 1 min at 94°C (denaturation). This profile was repeated for 25 cycles and the final 72°C extension was increased to 10 min. The aqueous layers of the reaction mixtures were pooled and extracted once with 1:1 phenol/chloroform and twice with chloroform. The DNA was ethanol-precipitated and resuspended in 20  $\mu$ l TE buffer [10 mM Tris.HCl, (pH 7.5), 1 mM EDTA]. The DNA was then cleaved with appropriate restriction enzymes to generate cohesive ends and gel-purified.

The construction of an over-producer clone was done in two steps. First, the PCR-generated DNA containing the *fokIM* gene was digested with *NcoI* and gel purified. It was then ligated into *NcoI*-cleaved and dephosphorylated pACYC184 and the recombinant DNA transfected into *E.coli* RB791 *i<sup>q</sup>* or RR1 cells made competent as described by Maniatis et al (Maniatis et al., Molecular Cloning. A laboratory manual Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982). After Tc selection, several

clones were picked and plasmid DNA was examined by restriction analysis for the presence of *fokIM* gene fragment in correct orientation to the chloramphenicol promoter of the vector (see figure 2). This plasmid expresses *FokI* methylase constitutively and then protects the host from chromosomal cleavage, when the *fokIR* gene is introduced into this host on a compatible plasmid. The plasmid DNA from these clones are therefore resistant to *FokI* digestion.

Second, the PCR-generated *fokIR* fragment was ligated into *Bam*HI-cleaved and dephosphorylated high expression vectors pRRS or pCB. pRRS possesses a *lac* UV5 promoter and pCB containing the strong *tac* promoter. In addition, these vectors contain the positive retroregulator stem-loop sequence derived from the crystal protein-encoding gene of *Bacillus Thuringiensis* downstream of the inserted *fokIR* gene. The recombinant DNA was transfected into competent *E.coli* RB791 *i*<sup>q</sup> [pACYC*fokIM*] or RR1[pACYC*fokIM*] cells. After Tc and Ap antibiotic selection, several clones were picked and plasmid DNA was examined by restriction analysis for *fokIR* gene fragment in correct orientation for expression from the vector promoters. These constructs were then examined for enzyme production.

To produce the enzyme, plasmid-containing RB791 *i*<sup>q</sup> or RR1 cells were grown at 37°C with shaking in 2x concentrated TY medium [1.6% tryptone, 1% yeast extract, 0.5% NaCl (pH 7.2)] supplemented with 20 µg Tc/ml (except for the pUC*fokIRM* plasmid) and 50 µg Ap/ml. IPTG was added to a concentration of 1 mM when the cell density reached O.D.<sub>600</sub> = 0.8.

The cells were incubated overnight (12 hr) with shaking. As is shown in Figure 2, both constructs yield FokI to a level of 5-8% of the total cellular protein.

5

### Examples III

#### Purification of FokI endonuclease

A simple three-step purification procedure was used to obtain electrophoretically homogeneous FokI endonuclease. RR1 [pACYCfokIM, pRRSfokIR] were  
10 grown in 6L of 2 x TY containing 20µg Tc/ml and 50 µg/Ap ml at 37°C to  $A_{600} = 0.8$ . and then induced overnight with 1 mM IPTG. The cells were harvested by centrifugation and then resuspended in 250 ml of buffer A [10 mM Tris.phosphate (pH 8.0), 7 mM 2-  
15 mercaptoethanol, 1 mM EDTA, 10% glycerol] containing 50 mM NaCl.

The cells were disrupted at maximum intensity on a Branson Sonicator for 1 hr at 4°C. The sonicated cells were centrifuged at 12,000 g for  
20 2 hr at 4°C. The supernatant was then diluted to 1L with buffer A containing 50 mM NaCl. The supernatant was loaded onto a 10 ml phosphocellulose (Whatman) column pre-equilibrated with buffer A containing 50 mM NaCl. The column was washed with  
25 50 ml of loading buffer and the protein was eluted with a 80-ml total gradient of 0.05M to 0.5M NaCl in buffer A. The fractions were monitored by  $A_{280}$  absorption and analyzed by electrophoresis on SDS (0.1%)-polyacrylamide (12%) gels (Laemmli, Nature,  
30 222:680-685, 1970). Proteins were stained with coomassie blue.

Restriction endonuclease activity of the fractions were assayed using pTZ19R as substrate.

The fractions containing *FokI* were pooled and fractionated with ammonium sulfate. The 50-70% ammonium sulfate fraction contained the *FokI* endonuclease. The precipitate was resuspended in 50 ml of buffer A containing 25 mM NaCl and loaded onto a DEAE column. *FokI* does not bind to DEAE while many contaminating proteins do. The flow-through was concentrated on a phosphocellulose column. Further purification was achieved using gel filtration (ACA 44) column. The *FokI* was purified to electrophoretic homogeneity using this procedure.

SDS (0.1%) polyacrylamide (12%) gel electrophoresis profiles of protein species present at each stage of purification are shown in Figure 3. The sequence of the first ten amino acids of the purified enzyme was determined by protein sequencing. The determined sequence was the same as that predicted from the nucleotide sequence. Crystals of this purified enzyme have also been grown using PEG 4000 as the precipitant. *FokI* endonuclease was purified further using Aca44 gel filtration column.

#### Example IV

##### Analysis of *FokIR* endonuclease by trypsin cleavage in the presence of DNA substrate.

Trypsin is a serine protease and it cleaves at the C-terminal side of lysine and arginine residues. This is a very useful enzyme to study the domain structure of proteins and enzymes. Trypsin digestion of *FokI* in the presence of its substrate, d-5'-CCTCTGGATGCTCTC-3' (SEQ ID NO: 10): 5'-GAGAGCATCCAGAGG-3' (SEQ ID NO: 11) was carried out with an oligonucleotide duplex to *FokI* molar

ratio of 2.5:1. *FokI* (200  $\mu$ g) was incubated with the oligonucleotide duplex in a volume 180  $\mu$ l containing 10 mM Tris.HCl, 50 mM NaCl, 10% glycerol and 10 mM  $MgCl_2$  at RT for 1 hr. Trypsin (20  $\mu$ l, 0.2 mg/ml) was added to the mixture. Aliquots (28  $\mu$ l) from the reaction mixture were removed at different time intervals and quenched with excess trypsin inhibitor, antipain. The tryptic fragments were purified by reversed-phase HPLC and their N-terminus sequence determined using an automatic protein sequenator from Applied Biosystems.

The time course of trypsin digestion of *FokI* endonuclease in the presence of 2.5 molar excess of oligonucleotide substrate and 10 mM  $MgCl_2$  is shown in Figure 4. At the 2.5 min time point only two major fragments other than the intact *FokI* were present, a 41 kDa fragment and a 25 kDa fragment. Upon further trypsin digestion, the 41 kDa fragment degraded into a 30 kDa fragment and 11 kDa fragment. The 25 kDa fragment appeared to be resistant to any further trypsin digestion. This fragment appeared to be less stable if the trypsin digestion of *FokI* - oligo complex was carried out in the absence of  $MgCl_2$ .

Only three major fragments (30 kDa, 25 kDa and 11 kDa) were present at the 160 min time point. Each of these fragments (41 kDa, 30 kDa, 25 kDa and 11 kDa) was purified by reversed-phase HPLC and their N-terminal amino acid sequence were determined (Table I). By comparing these N-terminal sequences to the predicted sequence of *FokI*, the 41 kDa and 25 kDa fragments were identified as N-terminal and C-

terminal fragments, respectively. In addition, the 30 kDa fragment was N-terminal.

Example V

5 Isolation of DNA binding tryptic fragments of FokI endonuclease using oligo dT-cellulose affinity column.

The DNA binding properties of the tryptic fragments were analyzed using an oligo dT-cellulose column. FokI (160 µg) was incubated with the 2.5 molar excess oligonucleotide duplex [d-5'-  
10 CCTCTGGATGCTCTC(A)<sub>15</sub>-3' (SEQ ID NO: 14):  
5'GAGAGCATCCAGAGG(A)<sub>15</sub>-3' (SEQ ID NO: 15)] in a volume of 90 µl containing 10 mM Tris.HCl (pH 8), 50 mM NaCl, 10% glycerol and 10 mM MgCl<sub>2</sub> at RT for 1 hr.  
15 Trypsin (10 µl, 0.2 mg/ml) was added to the solution to initiate digestion. The ratio of trypsin to FokI (by weight) was 1:80. Digestion was carried out for 10 min to obtain predominantly 41 kDa N-terminal fragment and 25 kDa C-terminal fragments in the  
20 reaction mixture. The reaction was quenched with large excess of antipain (10 µg) and diluted in loading buffer [10 mM.Tris HCl (pH 8.0), 1 mM EDTA and 100 mM MgCl<sub>2</sub>] to a final volume of 400 µl.

The solution was loaded onto a oligo dT-cellulose column (0.5 ml, Sigma, catalog #0-7751) pre-equilibrated with the loading buffer. The breakthrough was passed over the oligo dT-cellulose column six times. The column was washed with 5 ml of loading buffer and then eluted twice with 0.4 ml  
30 of 10 mM Tris.HCl (pH 8.0), 1 mM EDTA. These fractions contained the tryptic fragments that were bound to the oligonucleotide DNA substrate. The tryptic fragment bound to the oligo dT-cellulose

column was analyzed by SDS-polyacrylamide gel electrophoresis.

In a separate reaction, the trypsin digestion was carried out for 160 min to obtain  
5 predominantly the 30 kDa, 25 kDa and 11 kDa fragments in the reaction mixture.

Trypsin digestion of *FokI* endonuclease for 10 min yielded the 41 kDa N-terminal fragment and 25 kDa C-terminal fragments as the predominant species  
10 in the reaction mixture (Figure 5, Lane 3). When this mixture was passed over the oligo dT-cellulose column, only the 41 kDa N-terminal fragment is retained by the column suggesting that the DNA binding property of *FokI* endonuclease is in the N-  
15 terminal 2/3's of the enzyme. The 25 kDa fragment is not retained by the oligo dT-cellulose column.

Trypsin digestion of *FokI* - oligo complex for 160 min yielded predominantly the 30 kDa, 25 kDa and 11 kDa fragments (Figure 5, Lane 5). When this  
20 reaction mixture was passed over oligo dT-cellulose column, only the 30 kDa and 11 kDa fragments were retained. It appears these species together bind DNA and they arise from further degradation of 41 kDa N-terminal fragment. The 25 kDa fragment was  
25 not retained by oligo dT-cellulose column. It also did not bind to DEAE and thus could be purified by passage through a DEAE column and recovering it in the breakthrough volume.

*FokI* (390  $\mu$ g) was incubated with 2.5 molar  
30 excess of oligonucleotide duplex [d-5'-CTCTGGATGCTCTC-3 (SEQ ID NO: 10)':5'-GAGAGCATCCAGAGG-3' (SEQ ID NO: 11)] in a total volume of 170  $\mu$ l containing 10 mM Tris.HCl (pH 8), 50 mM



NaCl and 10% glycerol at RT for 1 hr. Digestion with trypsin (30  $\mu$ l; 0.2 mg/ml) in the absence of  $MgCl_2$  was for 10 min at RT to maximize the yield of the 41 kDa N-terminal fragment. The reaction was  
5 quenched with excess antipain (200  $\mu$ l). The tryptic digest was passed through a DEAE column. The 25 kDa of C-terminal fragment was recovered in the breakthrough volume. All the other tryptic fragments (41 kDa, 30 kDa and 11 kDa) were retained  
10 by the column and were eluted with 0.5M NaCl buffer (3 x 200  $\mu$ l). In a separate experiment, the trypsin digestion of *FokI* -oligo complex was done in presence of 10 mM  $MgCl_2$  at RT for 60 min to maximize the yield of 30 kDa and 11 kDa fragments. This  
15 purified fragment cleaved non-specifically both unmethylated DNA substrate (pTZ19R; Figure 6) and methylated DNA substrate (pACYC*fokIM*) in the presence of  $MgCl_2$ . These products are small, indicating that it is relatively non-specific in  
20 cleavage. The products were dephosphorylated using calf intestinal phosphatase and rephosphorylated using polynucleotide kinase and [ $\gamma$ - $^{32}P$ ] ATP. The  $^{32}P$ -labeled products were digested to mononucleotides using DNase I and snake venom  
25 phosphodiesterase. Analysis of the mononucleotides by PEI-cellulose chromatography indicates that the 25 kDa fragment cleaved preferentially phosphodiester bonds 5' to G>A>>T-C. The 25 kDa C-terminal fragment thus constitutes the cleavage  
30 domain of *FokI* endonuclease.

The 41 kDa N-terminal fragment - oligo complex was purified by agarose gel electrophoresis. *FokI* endonuclease (200  $\mu$ g) was incubated with 2.5

molar excess of oligonucleotide duplex, [d-5' -  
CCTCTGGATGCTCTC-3' (SEQ ID NO: 10): 5'-  
GAGAGCATCCAGAGG-3' (SEQ ID NO: 11)] in a volume of 180  
μl containing 10 mM Tris.HCl (pH 8.0), 50 mM NaCl  
5 and 10% glycerol at RT for 1 hr. Tracer amounts of  
32P-labeled oligonucleotide duplex was incorporated  
into the complex to monitor it during gel  
electrophoresis. Digestion with trypsin (20 μl; 0.2  
mg/ml) was for 12 min at RT to maximize the yield of  
10 the 41 kDa N-terminal fragment. The reaction was  
quenched with excess antipain. The 41 kDa N-  
terminal fragment - oligo complex was purified by  
agarose gel electrophoresis. The band corresponding  
to the complex was excised and recovered by  
15 electroelution in a dialysis bag  
(~ 600 μl). Analysis of the complex by SDS-PAGE  
revealed 41 kDa N-terminal fragment to be the major  
component. The 30 kDa N-terminal fragment and the  
11 kDa C-terminal fragment were present as minor  
20 components. These together appeared to bind DNA and  
co-migrate with the 41 kDa N-terminal fragment-oligo  
complex.

The binding specificity of the 41 kDa N-  
terminal fragment was determined using gel mobility  
25 shift assays.

#### Example VI

##### Gel Mobility shift assays

The specific oligos (d-5'-CCTCTGGATGCTCTC-  
3' (SEQ ID NO: 10) and d-5'-GAGAGCATCCAGAGG-3' (SEQ  
30 ID NO: 11)) were 5'-32P-labeled in a reaction  
mixture of 25 μl containing 40 mM Tris.HCl (pH 7.5),  
20 mM MgCl<sub>2</sub>, 50 mM NaCl, 10 mM DTT, 10 units of T4  
polynucleotide kinase (from New England Biolabs) and

20  $\mu\text{Ci}$  [ $\gamma$ - $^{32}\text{P}$ ] ATP (3000 Ci/mmol). The mixture was incubated at 37°C for 30 min. The kinase was inactivated by heating the reaction mixture to 70°C for 15 min. After addition of 200  $\mu\text{l}$  of water, the solution was passed through Sephadex G-25 (Superfine) column (Pharmacia) to remove the unreacted [ $\gamma$ - $^{32}\text{P}$ ] ATP. The final concentration of labeled single-strand oligos were 27  $\mu\text{M}$ .

The single-strands were then annealed to form the duplex in 10 mM Tris.HCl (pH 8.0), 50 mM NaCl to a concentration of 12  $\mu\text{M}$ . 1  $\mu\text{l}$  of the solution contained ~ 12 picomoles of oligo duplex and ~ 50 x 10<sup>3</sup>cpm. The non-specific oligos (d-5'-TAATTGATTCTTAA-3' (SEQ ID NO: 12) and d-5'-ATTAAGAATCAATT-3' (SEQ ID NO:13)) were labeled with [ $\gamma$ - $^{32}\text{P}$ ]ATP and polynucleotide kinase as described herein. The single-stranded oligos were annealed to yield the duplex at a concentration of 12 $\mu\text{M}$ . 1  $\mu\text{l}$  of the solution contained ~ 12 picomoles of oligo duplex and ~ 25 x 10<sup>3</sup>cpm. The non-specific oligos (d-5'-TAATTGATTCTTAA-3' (SEQ ID NO: 12) and d-5'-ATTAAGAATCAATT-3' (SEQ ID NO: 13)) were labeled with [ $\gamma$ - $^{32}\text{P}$ ] ATP and polynucleotide Kinase as described herein. The single-strand oligos were annealed to yield the duplex at a concentration of 12 $\mu\text{M}$ . 1  $\mu\text{l}$  of the solution contained 42 picomdes of oligo duplex and ~25x10<sup>3</sup> cpm.

10  $\mu\text{l}$  of 41 kDa N-terminal fragment-oligo complex (~ 2 pmoles) in 10 mM Tris.HCl, 50 mM NaCl and 10 mM MgCl<sub>2</sub> was incubated with 1  $\mu\text{l}$  of  $^{32}\text{p}$ -labeled specific oligonucleotide duplex (or  $^{32}\text{p}$ -labeled non-specific oligonucleotide duplex) at 37°C for 30 min and 120 min respectively. 5  $\mu\text{l}$  of 75%

glycerol was added to each sample and loaded on a 8% nondenaturing polyacrylamide gel. Electrophoresis was at 300 volts in TBE buffer until bromophenol blue moved ~ 6 cm from the top of the gel. The gel  
5 was dried and autoradiographed.

The complex readily exchanged  $^{32}\text{P}$ -labeled specific oligonucleotide duplex that contained the *FokI* recognition site as seen from the gel mobility shift assays (Figure 7). It did not, however,  
10 exchange the  $^{32}\text{P}$ -labeled non-specific oligonucleotide duplex that did not contain the *FokI* recognition site. These results indicate that all the information necessary for sequence-specific recognition of DNA are encoded within the 41 kDa N-  
15 terminal fragment of *FokI*.

#### Example VII

##### Analysis of *FokI* by trypsin cleavage in the absence of DNA substrate.

A time course of trypsin digestion of *FokI* endonuclease in the absence of the DNA substrate is shown in Figure 8. Initially, *FokI* cleaved into a  
20 58 kDa fragment and a 8 kDa fragment. The 58 kDa fragment did not bind DNA substrates and is not retained by the oligo dT-cellulose column. On  
25 further digestion, the 58 kDa fragment degraded into several intermediate tryptic fragments. However, the complete trypsin digestion yielded only 25 kDa fragments (appears as two overlapping bands).

Each of these species (58 kDa, 25 kDa and  
30 8 kDa) were purified by reversed phase HPLC and their amino terminal amino acid sequence determined (Table I). Comparison of the N-terminal sequences to the predicted *FokI* sequence revealed that the 8

kDa fragment to be N-terminal and the 58 kDa fragment to be C-terminal. This further supports the conclusion that N-terminus of *FokI* is responsible for the recognition domain. Sequencing the N-terminus of the 25 kDa fragments revealed the presence of two different components. A time course of trypsin digestion of *FokI* endonuclease in the presence of a non-specific DNA substrate yielded a profile similar to the one obtained when trypsin digestion of *FokI* is carried out in absence of any DNA substrate.

#### Example VIII

##### Cleavage specificity of the 25 kDa C-terminal tryptic fragment of *FokI*

The 25 kDa C-terminal tryptic fragment of *FokI* cleaved pTZ19R to small products indicating non-specific cleavage. The degradation products were dephosphorylated by calf intestinal phosphatase and  $^{32}\text{P}$ -labeled with the polynucleotide kinase and  $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ . The excess label was removed using a Sephadex G-25 (Superfine) column. The labeled products were then digested with 1 unit of pancreatic DNase I (Boehringer-Mannheim) in buffer containing 50 mM Tris.HCl (pH7.6), 10mM  $\text{MgCl}_2$  at 37°C for 1 hr. Then, 0.02 units of snake venom phosphodiesterase was added to the reaction mixture and digested at 37°C for 1 hr.

#### Example IX

##### Functional domains in *FokI* restriction endonuclease.

Analysis of functional domains of *FokI* (in the presence and absence of substrates) using trypsin was summarized in Figure 9. Binding of DNA substrate by *FokI* was accompanied by alteration in

the structure of the enzyme. This study supports that presence of two separate protein domains within this enzyme: one for sequence-specific recognition and the other for endonuclease activity. The results indicate that the recognition domain is at the N-terminus of the *FokI* endonuclease, while the cleavage domain is probably in the C-terminus third of the molecule.

Examples Relating to Construction  
of Insertion Mutants (X-XIV)

The complete nucleotide sequence of the *FokI* RM system has been published by various laboratories (Looney et al., Gene 80: 193-208, 1989 & Kita et al., J. Biol.Chem. 264: 5751-56, 1989). Experimental protocols for PCR are described, for example, in Skoglund et al., Gene 88:1-5, 1990 and in Bassing et al., Gene 113:83-88, 1992. The procedures for cell growth and purification of the mutant enzymes are similar to the ones used for the wild-type *FokI* (Li et al., Proc. Nat'l. Acad. Sci. USA 89:4275-79, 1992). Additional steps which include Sephadex G-75 gel filtration and Heparin-Sepharose CL-6B column chromatography were necessary to purify the mutant enzymes to homogeneity.

Example X

Mutagenesis of *SpeI* Site at Nucleotide 162 within the *fokIR* Gene

The two step PCR technique used to mutagenize one of the *SpeI* sites within the *fokIR* gene is described in Landt et al., Gene 96: 125-28, 1990. The three synthetic primers for this protocol include: 1) the mutagenic primer (5'-TCATAA TAGCAACTAATTCTTTTGGATCTT-3') (see SEQ ID NO:24)

containing one base mismatch within the *SpeI* site;  
2) the other primers each of which are flanked by  
restriction sites *ClaI* (5'-CCATCGATATAGCCTTTTATT-  
3') (see SEQ ID NO:25) and *XbaI* (5'-  
5 GCTCTAGAGGATCCGGAGGT-3') (see SEQ ID NO:26),  
respectively. An intermediate fragment was  
amplified using the *XbaI* primer and the mutagenic  
primer during the first step. The *ClaI* primer was  
then added to the intermediate for the second step  
10 PCR. The final 0.3 kb PCR product was digested with  
*XbaI/ClaI* to generate cohesive ends and gel-  
purified. The expression vector (pRRSfokIR) was  
cleaved with *XbaI/ClaI*. The large 4.2 kb fragment  
was then gel-purified and ligated to the PCR  
15 fragment. The recombinant DNA was transfected into  
competent *E. coli* RR1[pACYCfokIM] cells. After  
tetracycline and ampicillin antibiotic selection  
several clones were picked, and their plasmid DNA  
was examined by restriction analysis. The *SpeI* site  
20 mutation was confirmed by sequencing the plasmid DNA  
using Sanger's sequencing method (Sanger et al.  
Proc. Natl. Acad. Sci. USA 74: 5463-67, 1977).

#### Example XI

##### Construction of four (or seven) codon Insertion

##### 25 Mutants

The PCR-generated DNA containing a four  
(or seven) codon insertion was digested with a  
*SpeI/XmaI* and gel-purified. The plasmid, pRRSfokIR  
from Example X was cleaved with *SpeI/XmaI*, and the  
30 large 3.9 kb fragment was gel-purified and ligated  
to the PCR product. The recombinant DNA was  
transfected into competent RR1[pACYCfokIM] cells,  
and the desired clones identified as described in

Example X. The plasmids from these clones were isolated and sequenced to confirm the presence of the four (or seven) codon insertion within the *fokIR* gene.

5                   In particular, the construction of the mutants was performed as follows: (1) There are two *SpeI* sites at nucleotides 162 and 1152, respectively, within the *fokIR* gene sequence. The site at 1152 is located near the trypsin cleavage  
10                   site of *FokI* that separates the recognition and cleavage domains. In order to insert the four (or seven) codons around this region, the other *SpeI* site at 162 was mutagenized using a two step PCR technique (Landt et al. Gene 96:125-28, 1990).  
15                   Introduction of this *SpeI* site mutation in the *fokIR* gene does not affect the expression levels of the overproducer clones. (2) The insertion of four (or seven) codons was achieved using the PCR technique. The mutagenic primers used in the PCR amplification  
20                   are shown in Figure 11. Each primer has a 21 bp complementary sequence to the *fokIR* gene. The 5' end of these primers are flanked by *SpeI* sites. The codons for KSEL and KSELEEK repeats are incorporated between the *SpeI* site and the 21 bp complement.  
25                   Degenerate codons were used in these repeats to circumvent potential problems during PCR amplification. The other primer is complementary to the 3' end of the *fokIR* gene and is flanked by a *XmaI* site. The PCR-generated 0.6 kb fragments  
30                   containing the four (or seven) codon inserts digested with *SpeI/XmaI* and gel-purified. These fragments were substituted into the high expression vector pRRS*fokIR* to generate the mutants. Several



clones of each mutant identified and their DNA sequence confirmed by Sanger's dideoxy chain termination method (Sanger et al. Proc. Natl. Acad. Sci. USA 74.5463-67 1977).

5           Upon induction with 1 mM isopropyl  $\beta$ -D-thiogalactoside (IPTG), the expression of mutant enzymes in these clones became most prominent at 3 hrs as determined by SDS/PAGE. This was further supported by the assays for the enzyme activity.

10   The levels of expression of the mutant enzymes in these clones were much lower compared to the wild-type *FokI*. IPTG induction for longer times resulted in lower enzyme levels indicating that the mutant enzymes were actively degraded within these clones.

15   This suggests that the insertion of four (or seven) codons between the recognition and cleavage domains of *FokI* destabilizes the protein conformation making them more susceptible to degradation within the cells. SDS/PAGE profiles of the mutant enzymes are

20   shown in Figure 12.

#### Example XII

##### Preparation of DNA Substrates with a Single *FokI* Site

25           Two substrates, each containing a single *FokI* recognition site, were prepared by PCR using pTZ19R as the template. Oligonucleotide primers, 5'-CGCAGTGTTATCACTCAT-3' and 5'-CTTGTTGAGTACTCACC-3' (see SEQ ID NO:27 and SEQ ID NO:28, respectively), were used to synthesize the 100 bp fragment.

30   Primers, 5'-ACCGAGCTCGAATTCAT-3' and 5'-GATTTCCGGCCTATTGGTT-3' (see SEQ ID NO:29 and SEQ ID NO:30, respectively), were used to prepare the 256 bp fragment. Individual strands within these

substrates were radiolabeled by using the corresponding  $^{32}\text{P}$ -labeled phosphorylated primers during PCR. The products were purified from low-melting agarose gel, ethanol precipitated and resuspended in TE buffer.

#### Example XIII

##### Analysis of the Sequence Specificity of the Mutant Enzymes

The agarose gel electrophoretic profile of the cleavage products of pTZ19R DNA by FokI and the mutants are shown in Figure 13A. They are very similar suggesting that insertion of four (or seven) codons in the linker region between the recognition and cleavage domains does not alter its DNA sequence specificity. This was further confirmed by using  $^{32}\text{P}$ -labeled DNA substrates (100 bp and 256 bp) each containing a single FokI site. Substrates containing individual strands labeled with  $^{32}\text{P}$  were prepared as described in Example XII. FokI cleaves the 256 bp substrate into two fragments, 180 bp and 72 bp, respectively (Figure 13B). The length of the fragments was calculated from the  $^{32}\text{P}$ -labeled 5' end of each strand. The autoradiograph of the agarose gel is shown in Figure 13C. Depending on which strand carries the  $^{32}\text{P}$ -label in the substrate, either 72 bp fragment or 180 bp fragment appears as a band in the autoradiograph. The mutant enzymes reveal identical agarose gel profiles and autoradiograph. Therefore, insertion of four (or seven) codons between the recognition and cleavage domains does not alter the DNA recognition mechanism of FokI endonuclease.

Example XIVAnalysis of the Cleavage Distances from the  
Recognition Site by the Mutant Enzymes

To determine the distance of cleavage by  
5 the mutant enzymes, their cleavage products of the  
32P-labeled substrates were analyzed by PAGE (Figure  
14). The digests were analyzed alongside the  
sequencing reactions of pTZ19R performed with the  
same primers used in PCR to synthesize these  
10 substrates. The cleavage pattern of the 100 bp  
fragment by FokI and the mutants are shown in Figure  
14A. The cut sites are shifted from the recognition  
site on both strands of the substrates in the case  
of the mutants, as compared to the wild-type enzyme.  
15 The small observable shifts between the sequencing  
gel and the cleavage products are due to the  
unphosphorylated primers that were used in the  
sequencing reactions.

On the 5'-GGATG-3' strand, both mutants  
20 cut the DNA 10 nucleotides away from the site while  
on the 5'-CATCC-3' strand they cut 14 nucleotides  
away from the recognition site. These appear to be  
the major cut sites for both the mutants. A small  
amount of cleavage similar to the wild-type enzyme  
25 was is also observed.

The cleavage pattern of the 256 bp  
fragment is shown in Figure 14B. The pattern of  
cleavage is shown in Figure 14B. The pattern of  
cleavage is similar to the 100 bp fragment. Some  
30 cleavage is seen 15 nucleotides away from the  
recognition site on the 5'-CATCC-3' strand in the  
case of the mutants. The multiple cut sites for the  
mutant enzymes could be attributed to the presence

of different conformations in these proteins. Or  
 due to the increased flexibility of the spacer  
 region between the two domains. Depending on the  
 DNA substrate, some variation in the intensity of  
 5 cleavage at these sites was observed. This may be  
 due to the nucleotide sequence around these cut  
 sites. Naturally occurring Type IIS enzymes with  
 multiple cut sites have been reported (Szybalski et  
 al., Gene 100:13-26, 1991).

10

TABLE 1

Amino-terminal sequences of *FokI*  
 fragments from trypsin digestion

15	Fragment	Amino-terminal sequence	DNA substrate	SEQ ID NO
	8 kDa	VSKIRTFG*VQNPGKFENLKRVVQVFDRS	-	16
	58 kDa	SEAPCDAI IQ		17
	25 kDa	QLVKSELEEK	+	18
20	41 kDa	VSKIRTFGWV		19
	30 kDa	VSKIRTFGWV		19
	11 kDa	FTRVPKRVY		20

TABLE 2

No. (1)	ENase-IIs <sup>a</sup> (isoschi- zomers) (2)	Protruding ends <sup>c</sup> (5)	Species (strain) <sup>d</sup> (6)	Co-produced ENases <sup>c</sup> (7)	Described MTases-II <sup>f</sup> [C or A] (8)	Commercial availability (9)	References (10)
1.	AlwI (BinI) (BthII) <sup>1</sup>	5'N <sub>1</sub>	Acinetobacter lwofii			N, Z	Mo2, Ne3
2.	AlwXI (BbVI)	5'N <sub>4</sub>	Acinetobacter lwofii X		(M.BbVI)		Mo6
3.	Alw261 (BsmAI)	5'N <sub>4</sub>	Acinetobacter lwofii RFL26		M.Alw261 [C-5 and A-N6]		Gl1, Bl2
4.	BbsI (BbvII)	5'N <sub>4</sub>	Bacillus brevis (laterosporus NEB573)			N	Mo2, Ne3
5.	BbVI (AlwXI) (Uba11091) , (Bsp4321)	5'N <sub>4</sub>	Bacillus brevis (ATCC 9999)	BbvII	M.BbVI [C-5]	G, I, N, Z	Ba4, Do1, Do2, Gl2, Gl3, Ha4, Ha5, Ne3, SC2, Val
6.	BbvII (BbvI6I) <sup>1</sup> (BspVI)	5'N <sub>4</sub>	Bacillus brevis 80	BbVI			Bu1, Bu2, Do2, Ma4
7.	Bcefi	5'N <sub>1</sub>	Bacillus cereus subsp. flourescens				Ve1, Ve2
8.	BcoI		Bacteroides cacciae			(N)	Mo2

No. (1)	ENase-III <sup>a</sup> (isocho- mers) (2)	Protruding ends <sup>c</sup> (5)	Species (strain) <sup>d</sup> (6)	Co-produced ENases <sup>c</sup> (7)	Described MTases-II <sup>f</sup> [C or A] (8)	Commercial availability (9)	References (10)
9.	BcgI	3'N <sub>2</sub> 3'N <sub>2</sub>	Bacillus coagulans (NEB 566)			N	H. Kong, No3
10.	BlnI (AlwI) (BthII)	5'N <sub>1</sub>	bifidobacter- ium infantis			N	Bo2, Kh1, Kh2
11.	BstI (Eco31I) <sup>j</sup>	5'N <sub>4</sub>	Bacillus stearothermo- philus 6-55			N	H. Kong, Mo2, Ne3
12.	BsgI	3'N <sub>2</sub>	Bacillus sphaericus GC			N	Sc2
13.	BsmAI (Alw26I)	5'N <sub>4</sub>	Bacillus stearothermo- philus A664 (NEB 481)			N	Ch1, Kol, Ne3
14.	BspMI	5'N <sub>4</sub>	Bacillus species M (NEB 356)	BspMII		N	Ha1, Ki2, Ki4, Ku1, Mc2, Mo2, Mo4, Mo7 Ne3
15.	EarI (Ksp632I)	5'N <sub>3</sub>	Enterobacter aerogenes (NEB 450)			N	Ne3, Po3
16.	Eco31I (BsaI) <sup>j</sup>	5'N <sub>4</sub>	Escherichia coli RFL31		M.Eco31I [C-5] and [A-N6]	F	Bl2, Bu3

No. (1)	ENase-II <sup>a</sup> (isocho- mers) (2)	Protruding ends (5)	Species (strain) <sup>d</sup> (6)	Co-produced ENases <sup>c</sup> (7)	Described MTases-II <sup>b</sup> [C or A] (8)	Commercial availability (9)	References (10)
17.	Eco57I (Bsp6II) <sup>1</sup> {Eco112I} {Eco125I} (FsfI) <sup>1</sup>	3'N <sub>2</sub>	Escherichia coli RFL57,		M.Eco57I [A- N6]	F,N	Ja2, Ja3, Pe1, Pe2
18.	Esp3I	5'N <sub>4</sub>	Erwinia sp RFL3		M.Esp3I [C-5, A-N6]	F,N	B12
19.	Fai	5'N <sub>2</sub>	Flavobacter- ium aquatili				B12
20.	FokI (HincII)	5'N <sub>4</sub>	Flavobacte- rium oceanokoites		M.FokI [A-N6]	A,M,N,S,U,Z	Ba4, Ha2, Ha3, Ka1, Ka2, Ki1, Ki3, Ki4, Ki5, Ki6, Ki7, Kr1, La1, Lo1, Lu1, Ma1, Ma3, Mc1, Ne3, Nw1, Po1, Po4, Po5, Po6, Sc3, sc4, sk1, su2, Su3, Su4, Sz1, Ve3, Ve4, W11

No. (1)	ENase-II <sup>a</sup> (isoschizomers) (2)	Protruding ends (5)	Species (strain) <sup>d</sup> (6)	Co-produced ENases <sup>c</sup> (7)	Described MTases-II <sup>b</sup> [C or A] (8)	Commercial availability (9)	References (10)
21.	GsuI (Bco35I) 1 (Bsp22I) 1 (Bsp28I) 1	3'N <sub>2</sub>	Glucobacter dioxycace- tonicus H015T		M.GsuI	P, N	B11, Jal, Pel, Pe2
22.	HgaI	5'N <sub>5</sub>	Haemophilus gallinarum (ATCC14385)		M.HgaI (two MTases) [C-5]	N, Z	Ba4, Br1, Br6, Ko4, Krl, Mo8, Nel, Ne3, Sul, Tal, Tol, Uri
23.	HinGII (FokI)	5'N <sub>4</sub>	Haemophilus influenzae GU				Na2
24.	HphI (NgovII) (NgobI)	3'N <sub>1</sub> (or blunt)	Haemophilus parahaemoly- ticus		M.HphI [A-N6]	N, Z	Ba2, Col, Kl1, Ne2, Ne3, Ro1
25.	Ksp632I (EarI) 1 (BsrEI) 1	5'N <sub>3</sub>	Kluyvera sp. 632			M	Bo1
26.	MboII (NcuI) 1 (TceI) 1	3'N <sub>1</sub>	Moraxella bovis (ATCC10900)	MboI	M.MboII [A-N6]	B, G, I, N, P, U, Z	Ba1, Br3, Br5, En1, Gal, Gel, Ha2, Mc1, Mc3, Na1, Na2, Ne2, Ne3, Sc1, Se1 Sm1
27.	MmeI	3'N <sub>2</sub>	Methylophilus methylophilus	MmeII		U	Bo3, Tu1



No. (1)	ENase-IIIs <sup>a</sup> (isochi- zomers) (2)	Protruding ends <sup>c</sup> (5)	Species (strain) <sup>d</sup> (6)	Co-produced ENases <sup>c</sup> (7)	Described MTases-II <sup>f</sup> [C or A] (8)	Commercial availability (9)	References (10)
28.	MnlI	3'N <sub>1</sub>	Moraxella nonliquefa- ciens (ATCC17953)			I, N, S, Z	Br2, Ne3, Sc2, V11, Ea1
29.	NgovIII (HphI)	n.d.	Neisseria gonorrhoeae		M.NgovVIII		Ko2
30.	PleI	5'N <sub>1</sub>	Pseudomonas	lemoigne1 (NEB418)		N	Mo6, Ne3
31.	RleAI	3'N <sub>3</sub>	Rhizobium leguminosarum				Ve5
32.	SapI	5'N <sub>3</sub>	Saccharo- polyspora sp.			N	Mo2, Ne3
33.	SfaNI (BscAI) <sup>1</sup>	5'N <sub>4</sub>	Streptococcus faecalis ND547		M.SfaI	N, Z	Ba4, Ne3, Po5, Po6, Sc2, Sc3, Sc5, Sp1
34.	TaqII	3'N <sub>2</sub>	Thermus aquaticus	TaqI		U	Ba2, My1
35.	Tth1111I	3'N <sub>2</sub>	Thermus thermophilus 111	Tth111I		Y, Z	Sh1, Sh2
36. Related ENases: h	Sts I	Streptococ- cus sanguis 54					

No. (1)	ENase-IIs <sup>a</sup> (isochi- zomers) (2)	Protruding ends <sup>c</sup> (5)	Species <sup>d</sup> (strain) <sup>d</sup> (6)	Co-produced ENases <sup>c</sup> (7)	Described MTases-II <sup>b</sup> [C or A] (8)	Commercial availability (9)	References (10)
36.	BsmI (Asp35HI) X	3'N <sub>1</sub>	Bacillus stearo- thermophilus NUB36			N	Gil, Ha6, Inl, MO7, My1, Ne3, Pal
37.	BarI (BarSI)	3'N <sub>1</sub>	Bacillus stearothermo- philus (NEB447)			N	Ne3, Po2

<sup>a</sup> Class-II restriction endonucleases (ENases-IIS) as listed (Kel:Ro2). Isochizomers are listed in parentheses (very recently discovered or incompletely characterized isochizomers are in footnotes i-k). An ENase-IIS is defined as an enzyme which cuts at precise distance away from its recognition site, without cleaving this site. Enzymes in lines 36 and 37 (BsmI, BsrI, six AspI and BsgCI) do not fit this definition because one of the two cuts is within the recognition site, but they were included because several of their properties and applications are quite similar to those of enzymes 1-35. ENase in line 29 (NgovII) was not described, but the M.Ngo VIII MTase appears to match the HphI. Genes coding for Eco57I and FokI were cloned (Jal:Wil). ENases EcoI, Eco57I and GsuI (and their isochizomers?) require or are stimulated by AdoMet.

<sup>b</sup> The recognition sequences are asymmetric [with exception of those marked S (in bp column) which display a partial symmetry (which might be incidental)], and are oriented so that the cut sites are to the right (downstream) of them. E.g., GGATC(N)<sub>4</sub> (line 1), indicates that the cut on the upper strand is between 4th CCTAG(N)<sub>5</sub>

and 5th nt beyond C; on the lower strand the cut is between 5th and 6th nt beyond G. Length of the recognition site is given in bp, and the symbols + or - below it indicate whether the purified enzyme cuts (+) or does not (-) ss DNA. N, A, or G or T; R, A or G, C or T.

c As deduced from cut sites (see column 3). n.d., not determined.

d Strains which produce the specified Enases-IIS.

e Other unrelated Enases produced by the same strain.

f MTases-IIS isolated from the same strain. Genes bbvIM, eco57IM, fokIM, hgaIM, mbolIM and sfaIM (coding for M.BbvI, M.Eco57I, M.FokI, M.HgaI, M.MbolI, and M.SfaI, respectively; Sz3) were cloned (Ba4; BoO; Ja3; Wil). MTases with the same site specificity, but produced by another strain, are in parentheses. Methylated based (m<sup>5</sup>C or mN<sup>6</sup>A) as shown in brackets (as C-5 or A-N6, respectively).

g A, Amersham Buchler, Buckinghamshire (U.K.); B, BRL/Life Technologies, Gaithersburg, MD; F, ESP Fermentas, 2328 Vilnius, Lithuania (U.S.S.R.) (some also available from N); G, Anglian Biotechnology, Colchester (u.k.); 1, IBI/International Biotechnology, New Haven, CT; M, Boehringer/Mannheim, Mannheim (F.R.G.); N, New England Biolabs, Beverly, MA; P, PL-Pharmacia, Milwaukee, WI; S, Stratagene, La Jolla, Ca; U, Dept. of Microbiology, University of Gdansk, Gdansk (Poland); Y, NY Biolabs, New York, NY; Z, see American Chemical Society Biotech buyers' Guide (1991). Parentheses indicate that the ENase is produced, but not yet commercially available.

h These enzymes do not formally belong to class IIS (see footnotes). They are also designated IIT (Kel); (N)-1 indicates a cut within the recognition site in the lower strand (see arrowhead).

- i Cuts unknown (See Ro2).
  - j Also 28 additional ENases: Cfr561, Eco42 Eco511, Eco951, Eco971, Eco1011, Eco1201, Eco1271, Eco1291, Eco1551, Eco1561, Eco1571, Eco1621, 1851 Eco1911 Eco2031, Eco2051, Eco2171, Eco2251, Eco2391, Eco2401, Eco2411, Eco2461, Eco2471, pal, Sau121, which have the same recognition sequence, but for most of them cuts are unknown (see Ro2). PpaI has the same cut as Eco311 (ne31).
  - k Also additional isochizomers Asp26HI, Asp27HI, Asp36HI Asp40HI, Asp50HI (Ro21), and BacCI (from Bacillus sp. 2G).
- from Szybalski et al. [Gene 100:14-26 (1991)].

51

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

(i) APPLICANT: Chandrasegaran, Srinivasan

(ii) TITLE OF INVENTION: Functional Domains in  
FokI Restriction Endonuclease

(iii) NUMBER OF SEQUENCES: 40

## (iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Cushman, Darby & Cushman  
(B) STREET: 1100 New York Ave., N.W.  
(C) CITY: Washington  
(D) STATE: D.C.  
(E) COUNTRY: USA  
(F) ZIP: 20005-3918

## (v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version

#1.25

## (vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:  
(B) FILING DATE:  
(C) CLASSIFICATION:

## (viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Kokulis, Paul N.  
(B) REGISTRATION NUMBER: 16,773  
(C) REFERENCE/DOCKET NUMBER:  
PNK/4130/122364/CLB

## (ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 202-861-3503  
(B) TELEFAX: 202-822-0944  
(C) TELEX: 6714627 CUSH

52

## (2) INFORMATION FOR SEQ ID NO:1:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:

GGATG

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:

CCTAC

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 18..35

53

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CCATGGAGGT TTAAAT ATG AGA TTT ATT GGC AGC 35  
Met Arg Phe Ile Gly Ser  
1 5

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Arg Phe Ile Gly Ser  
1 5

## (2) INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ATACCATGGG AATTAAATGA CACAGCATCA 30

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

54

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 22..42

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
AGGATCCGG AGGTTTAAAA T ATG GTT TCT AAA ATA AGA ACT    42
               Met Val Ser Lys Ile Arg Thr
                1                      5
```

## (2) INFORMATION FOR SEQ ID NO:7:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
Met Val Ser Lys Ile Arg Thr
  1                      5
```

## (2) INFORMATION FOR SEQ ID NO:8:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
TAGGATCCTC ATTAAAAGTT TATCTGCCCG TTATT    35
```



55

## (2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 7 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Asn Asn Gly Glu Ile Asn Phe  
1 5

## (2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 15 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CCTCTGGATG CTCTC

15

## (2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 15 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GAGAGCATCC AGAGG

15

56

## (2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 14 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TAATTGATTC TTAA

14

## (2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 14 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

ATTAAGAATC AATT

14

## (2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 30 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CCTCTGGATG CTCTCAAAAA AAAAAAAAAA

30

57

## (2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 30 base pairs  
    (B) TYPE: nucleic acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GAGAGCATCC AGAGGAAAAA AAAAAAAAAA

30

## (2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 29 amino acids  
    (B) TYPE: amino acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Val	Ser	Lys	Ile	Arg	Thr	Phe	Gly	Xaa	Val	Gln	Asn	Pro	Gly
1			5				10						
Lys	Phe	Glu	Asn	Leu	Lys	Arg	Val	Val	Gln	Val	Phe	Asp	Arg
15				20					25				
Ser													

## (2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 10 amino acids  
    (B) TYPE: amino acid  
    (C) STRANDEDNESS: single  
    (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

58

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Ser Glu Ala Pro Cys Asp Ala Ile Ile Gln  
1 5 10

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 10 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Gln Leu Val Lys Ser Glu Leu Glu Glu Lys  
1 5 10

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 10 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Val Ser Lys Ile Arg Thr Phe Gly Trp Val  
1 5 10

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 9 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

59

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Phe Thr Arg Val Pro Lys Arg Val Tyr  
1 5

## (2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 3 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Glu Glu Lys  
1

## (2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 4 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Lys Ser Glu Leu  
1

## (2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Lys Ser Glu Leu Glu Glu Lys  
1 5

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

TAGCAACTAA TTCTTTTGG ATCTT

25

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CCATCGATAT AGCCTTTTTT ATT

23

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

61

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GCTCTAGAGG ATCCGGAGGT

20

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

CGCAGTGTTA TCACTCAT

18

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CTTGTTGAG TACTCACC

18

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

62

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

ACCGAGCTCG AATTCAC

18

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GATTTTCGGCC TATTGGTT

18

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 579 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Met	Val	Ser	Lys	Ile	Arg	Thr	Phe	Gly	Trp	Val	Gln	Asn	Pro
1				5					10				
Gly	Lys	Phe	Glu	Asn	Leu	Lys	Arg	Val	Val	Gln	Val	Phe	Asp
15				20					25				
Arg	Asn	Ser	Lys	Val	His	Asn	Glu	Val	Lys	Asn	Ile	Lys	Ile
30				35					40				
Pro	Thr	Leu	Val	Lys	Glu	Ser	Lys	Ile	Gln	Lys	Glu	Leu	Val
	45					50					55		
Ala	Ile	Met	Asn	Gln	His	Asp	Leu	Ile	Tyr	Thr	Tyr	Lys	Glu
	60					65						70	
Leu	Val	Gly	Thr	Gly	Thr	Ser	Ile	Arg	Ser	Glu	Ala	Pro	Cys
	75					80							
Asp	Ala	Ile	Ile	Gln	Ala	Thr	Ile	Ala	Asp	Gln	Gly	Asn	Lys
85				90					95				
Lys	Gly	Tyr	Ile	Asp	Asn	Trp	Ser	Ser	Asp	Gly	Phe	Leu	Arg
100					105						110		



63

Trp	Ala	His	Ala	Leu	Gly	Phe	Ile	Glu	Tyr	Ile	Asn	Lys	Ser
		115					120					125	
Asp	Ser	Phe	Val	Ile	Thr	Asp	Val	Gly	Leu	Ala	Tyr	Ser	Lys
		130						135					140
Ser	Ala	Asp	Gly	Ser	Ala	Ile	Glu	Lys	Glu	Ile	Leu	Ile	Glu
			145						150				
Ala	Ile	Ser	Ser	Tyr	Pro	Pro	Ala	Ile	Arg	Ile	Leu	Thr	Leu
155				160						165			
Leu	Glu	Asp	Gly	Gln	His	Leu	Thr	Lys	Phe	Asp	Leu	Gly	Lys
	170				175						180		
Asn	Leu	Gly	Phe	Ser	Gly	Glu	Ser	Gly	Phe	Thr	Ser	Leu	Pro
		185				190						195	
Glu	Gly	Ile	Leu	Leu	Asp	Thr	Leu	Ala	Asn	Ala	Met	Pro	Lys
			200					205					210
Asp	Lys	Gly	Glu	Ile	Arg	Asn	Asn	Trp	Glu	Gly	Ser	Ser	Asp
			215						220				
Lys	Tyr	Ala	Arg	Met	Ile	Gly	Gly	Trp	Leu	Asp	Lys	Leu	Gly
225				230						235			
Leu	Val	Lys	Gln	Gly	Lys	Lys	Glu	Phe	Ile	Ile	Pro	Thr	Leu
	240				245						250		
Gly	Lys	Pro	Asp	Asn	Lys	Glu	Phe	Ile	Ser	His	Ala	Phe	Lys
		255				260						265	
Ile	Thr	Gly	Glu	Gly	Leu	Lys	Val	Leu	Arg	Arg	Ala	Lys	Gly
			270					275					280
Ser	Thr	Lys	Phe	Thr	Arg	Val	Pro	Lys	Arg	Val	Tyr	Trp	Glu
			285						290				
Met	Leu	Ala	Thr	Asn	Leu	Thr	Asp	Lys	Glu	Tyr	Val	Arg	Thr
295				300						305			
Arg	Arg	Ala	Leu	Ile	Leu	Glu	Ile	Leu	Ile	Lys	Ala	Gly	Ser
	310				315						320		
Leu	Lys	Ile	Glu	Gln	Ile	Gln	Asp	Asn	Leu	Lys	Lys	Leu	Gly
		325				330						335	
Phe	Asp	Glu	Val	Ile	Glu	Thr	Ile	Glu	Asn	Asp	Ile	Lys	Gly
			340					345					350
Leu	Ile	Asn	Thr	Gly	Ile	Phe	Ile	Glu	Ile	Lys	Gly	Arg	Phe
			355						360				
Tyr	Gln	Leu	Lys	Asp	His	Ile	Leu	Gln	Phe	Val	Ile	Pro	Asn
365				370						375			
Arg	Gly	Val	Thr	Lys	Gln	Leu	Val	Lys	Ser	Glu	Leu	Glu	Glu
	380				385						390		
Lys	Lys	Ser	Glu	Leu	Arg	His	Lys	Leu	Lys	Tyr	Val	Pro	His
		395				400						405	
Glu	Tyr	Ile	Glu	Leu	Ile	Glu	Ile	Ala	Arg	Asn	Ser	Thr	Gln
			410					415					420
Asp	Arg	Ile	Leu	Glu	Met	Lys	Val	Met	Glu	Phe	Phe	Met	Lys
			425						430				
Val	Tyr	Gly	Tyr	Arg	Gly	Lys	His	Leu	Gly	Gly	Ser	Arg	Lys
435				440						445			
Pro	Asp	Gly	Ala	Ile	Tyr	Thr	Val	Gly	Ser	Pro	Ile	Asp	Tyr
	450					455					460		

64

Gly	Val	Ile	Val	Asp	Thr	Lys	Ala	Tyr	Ser	Gly	Gly	Tyr	Asn
	465						470					475	
Leu	Pro	Ile	Gly	Gln	Ala	Asp	Glu	Met	Gln	Arg	Tyr	Val	Glu
		480						485					490
Glu	Asn	Gln	Thr	Arg	Asn	Lys	His	Ile	Asn	Pro	Asn	Glu	Trp
			495						500				
Trp	Lys	Val	Tyr	Pro	Ser	Ser	Val	Thr	Glu	Phe	Lys	Phe	Leu
505					510					515			
Phe	Val	Ser	Gly	His	Phe	Lys	Gly	Asn	Tyr	Lys	Ala	Gln	Leu
	520					525					530		
Thr	Arg	Leu	Asn	His	Ile	Thr	Asn	Cys	Asn	Gly	Ala	Val	Leu
		535					540					545	
Ser	Val	Glu	Glu	Leu	Leu	Ile	Gly	Gly	Glu	Met	Ile	Lys	Ala
			550					555					560
Gly	Thr	Leu	Thr	Leu	Glu	Glu	Val	Arg	Arg	Lys	Phe	Asn	Asn
				565					570				
Gly	Glu	Ile	Asn	Phe									
575													

## (2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 11 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Lys	Gln	Leu	Val	Lys	Ser	Glu	Leu	Glu	Glu	Lys
1				5					10	

## (2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 33 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

65

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

AAGCAACTAG TCAAAAGTGA ACTGGAGGAG AAG

33

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Leu	Val	Lys	Ser	Glu	Leu	Lys	Ser	Glu	Leu	Glu	Glu	Lys
1				5					10			

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GGACTAGTCA AATCTGAACT TAAAAGTGAA CTGGAGGAGA AG

42

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

66

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Leu Val Lys Ser Glu Leu Glu Glu Lys Lys Ser Glu Leu Glu  
1 5 10  
Glu Lys  
15

## (2) INFORMATION FOR SEQ ID NO:37:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

GGACTAGTCA AATCTGAACT TGAGGAGAAG AAAAGTGAAC  
TGGAGGAGAA G

51

## (2) INFORMATION FOR SEQ ID NO:38:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Asn Phe Xaa Xaa  
1

## (2) INFORMATION FOR SEQ ID NO:39:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

67

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

TTGAAAATTA CTCCTAGGGG CCCCCCT

27

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GGATGNNNNNNNNNNNNNNNNNNNN

23

\* \* \* \* \*

All publications mentioned hereinabove are hereby incorporated by reference.

While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be appreciated by one skilled in the art that various changes in form and detail can be made without departing from the true scope of the invention.

WHAT IS CLAIMED IS:

1. An isolated DNA segment encoding the recognition domain of a Type IIS endonuclease which contains the sequence-specific recognition activity of said Type IIS endonuclease.

2. The DNA segment of claim 1 wherein said Type IIS endonuclease is *FokI* restriction endonuclease.

3. The DNA segment of claim 2 which encodes amino acids 1-382 of the *FokI* restriction endonuclease.

4. An isolated DNA segment encoding the catalytic domain of a Type IIS endonuclease which contains the cleavage activity of said Type IIS endonuclease.

5. The DNA segment of claim 4 wherein said Type IIS endonuclease is *FokI* restriction endonuclease.

6. The DNA segment of claim 5 which encodes amino acids 383-578 of the *FokI* restriction endonuclease.

7. An isolated protein consisting essentially of the N-terminus of the *FokI* restriction endonuclease which protein has the sequence-specific recognition activity of said endonuclease.

8. An isolated protein consisting essentially of the C-terminus of the *FokI* restriction endonuclease which protein has the cleavage activity of said endonuclease.

9. A DNA construct comprising:

(i) a first DNA segment encoding the catalytic domain of a Type IIS endonuclease which contains the cleavage activity of said Type IIS endonuclease;

(ii) a second DNA segment encoding a sequence-specific recognition domain other than the recognition domain of said Type IIS endonuclease; and

(iii) a vector

wherein said first DNA segment and said second DNA segment are operably linked to said vector so that a single protein is produced.

10. The DNA construct according to claim 9 wherein said Type IIS endonuclease is *FokI* restriction endonuclease.

11. The DNA construct according to claim 10 wherein said recognition domain is selected from the group consisting of: zinc finger motifs, homeo domain motifs, DNA binding domains of repressors, DNA binding domains of oncogenes and naturally occurring sequence-specific DNA binding proteins that recognize >6 base pairs.



12. A procaryotic cell comprising:

(i) a first DNA segment encoding the catalytic domain of a Type IIS endonuclease which contains the cleavage activity of said Type IIS endonuclease;

(ii) a second DNA segment encoding a sequence-specific recognition domain other than the recognition domain of said Type IIS endonuclease; and

(iii) a vector

wherein said first DNA segment and said second DNA segment are operably linked to said vector so that a single protein is produced.

13. A hybrid restriction enzyme comprising the catalytic domain of a Type IIS endonuclease which contains the cleavage activity of said Type IIS endonuclease covalently linked to a recognition domain of a protein other than said Type IIS endonuclease.

14. The hybrid restriction enzyme of claim 13 wherein said recognition domain which comprises part of said hybrid restriction enzyme is selected from the group consisting of: zinc finger motifs, homeo domain motifs, DNA binding domains of repressors, DNA binding domains of oncogenes and naturally occurring sequence-specific DNA binding proteins that recognize >6 base pairs.

15. A DNA construct comprising:

(i) a first DNA segment encoding the catalytic domain of a Type IIS endonuclease which

contains the cleavage activity of said Type IIS endonuclease;

(ii) a second DNA segment encoding a sequence-specific recognition domain other than the recognition domain of said Type IIS endonuclease;

(iii) a third DNA segment comprising one or more codons, wherein said third DNA segment is inserted between said first DNA segment and said second DNA segment; and

(iv) a vector

wherein said first DNA segment, said second DNA segment and said third DNA segment are operably linked to said vector so that a single protein is produced.

16. The DNA construct according to claim 15 wherein said Type IIS endonuclease is *FokI* restriction endonuclease.

17. The DNA construct according to claim 16 wherein said third DNA segment consists essentially of four codons.

18. The DNA construct according to claim 17 wherein said four codons of said third DNA segment are inserted at nucleotide 1152 of the gene encoding said endonuclease.

19. The DNA construct according to claim 16 wherein said third DNA segment consists essentially of 7 codons.

20. The DNA construct according to claim 19 wherein said 7 codons of said third DNA segment are inserted at nucleotide 1152 of the gene encoding said endonuclease.

21. The DNA construct according to claim 16 wherein said recognition domain is selected from the group consisting of: zinc finger motifs, homeo domain motifs, DNA binding domains of repressors, DNA binding domains of oncogenes and naturally occurring sequence-specific DNA binding proteins that recognize >6 base pairs.

22. A procaryotic cell comprising:

(i) a first DNA segment encoding the catalytic domain of a Type IIS endonuclease which contains the cleavage activity of said Type IIS endonuclease;

(ii) a second DNA segment encoding a sequence-specific recognition domain other than the recognition domain of said Type IIS endonuclease;

(iii) a third DNA segment comprising one or more codons, wherein said third DNA segment is inserted between said first DNA segment and said second DNA segment; and

(iv) a vector

wherein said first DNA segment, said second DNA segment, and said third DNA segment are operably linked to said vector so that a single protein is produced.

23. The procaryotic cell of claim 22 wherein said third DNA segment consists essentially of four codons.

24. The procaryotic cell of claim 22 wherein said third DNA segment consists essentially of seven codons.

25. An isolated protein produced by the procaryotic cell of claim 22.

26. An isolated DNA segment encoding the N-terminus of a Type IIS endonuclease which contains the sequence-specific recognition activity of said Type II endonuclease, said Type II endonuclease being *FokI* restriction endonuclease and having a molecular weight of about 41 kilodaltons as measured by SDS-polyacrylamide gel electrophoresis.

27. An isolated DNA segment encoding the C-terminus of a Type IIS endonuclease which contains the cleavage activity of said Type IIS endonuclease, said Type II endonuclease being *FokI* restriction endonuclease and having a molecular weight of about 25 kilodaltons as determined by SDS-polyacrylamide gel electrophoresis.

28. An isolated protein consisting essentially of the N-terminus of the *Fok* restriction endonuclease which protein has the sequence-specific recognition activity of said endonuclease and which protein is amino acids 1-382 of said *Fok* restriction endonuclease.

29. An isolated protein consisting essentially of the C-terminus of the *FokI* restriction endonuclease which protein has the nuclease activity of said endonuclease and which protein is amino acids 383-578 of said *FokI* restriction endonuclease.

## AMENDED CLAIMS

[received by the International Bureau  
on 11 July 1994 (11.07.94); original claims 1-8 replaced by new  
claims 1 and 2; original claims 9-25 and 28,29 renumbered  
as new claims 3-21 and 22,23; original claims  
26 and 27 cancelled (6 pages)]

1. An isolated DNA segment encoding the N-terminus of a Type IIS endonuclease which contains the sequence-specific recognition activity of said Type IIS endonuclease, said Type IIS endonuclease being *FokI* restriction endonuclease and said N-terminus having a molecular weight of about 41 kilodaltons as determined by SDS-polyacrylamide gel electrophoresis wherein said isolated DNA segment encodes amino acids 1-382 of said *FokI* restriction endonuclease.

2. An isolated DNA segment encoding the C-terminus of a Type IIS endonuclease which contains the cleavage activity of said Type IIS endonuclease, said Type IIS endonuclease being *FokI* and said C-terminus having a molecular weight of about 25 kilodaltons, as determined by SDS-polyacrylamide gel electrophoresis, wherein said isolated DNA segment encodes amino acids 383-578 of said *FokI* restriction endonuclease.

3. A DNA construct comprising:

(i) a first DNA segment encoding the catalytic domain of a Type IIS endonuclease which contains the cleavage activity of said Type IIS endonuclease;

(ii) a second DNA segment encoding a sequence-specific recognition domain other than the recognition domain of said Type IIS endonuclease;  
and

(iii) a vector

wherein said first DNA segment and said second DNA segment are operably linked to said vector so that a single protein is produced.

4. The DNA construct according to claim 3 wherein said Type IIS endonuclease is *FokI* restriction endonuclease.

5. The DNA construct according to claim 4 wherein said recognition domain is selected from the group consisting of: zinc finger motifs, homeo domain motifs, DNA binding domains of repressors, DNA binding domains of oncogenes and naturally occurring sequence-specific DNA binding proteins that recognize >6 base pairs.

6. A procaryotic cell comprising:

(i) a first DNA segment encoding the catalytic domain of a Type IIS endonuclease which contains the cleavage activity of said Type IIS endonuclease;

(ii) a second DNA segment encoding a sequence-specific recognition domain other than the recognition domain of said Type IIS endonuclease; and

(iii) a vector

wherein said first DNA segment and said second DNA segment are operably linked to said vector so that a single protein is produced.

7. A hybrid restriction enzyme comprising the catalytic domain of a Type IIS endonuclease which contains the cleavage activity of said Type

IIS endonuclease covalently linked to a recognition domain of a protein other than said Type IIS endonuclease.

8. The hybrid restriction enzyme of claim 7 wherein said recognition domain which comprises part of said hybrid restriction enzyme is selected from the group consisting of: zinc finger motifs, homeo domain motifs, DNA binding domains of repressors, DNA binding domains of oncogenes and naturally occurring sequence-specific DNA binding proteins that recognize >6 base pairs.

9. A DNA construct comprising:

(i) a first DNA segment encoding the catalytic domain of a Type IIS endonuclease which contains the cleavage activity of said Type IIS endonuclease;

(ii) a second DNA segment encoding a sequence-specific recognition domain other than the recognition domain of said Type IIS endonuclease;

(iii) a third DNA segment comprising one or more codons, wherein said third DNA segment is inserted between said first DNA segment and said second DNA segment; and

(iv) a vector

wherein said first DNA segment, said second DNA segment and said third DNA segment are operably linked to said vector so that a single protein is produced.



10. The DNA construct according to claim 9 wherein said Type IIS endonuclease is *FokI* restriction endonuclease.

11. The DNA construct according to claim 10 wherein said third DNA segment consists essentially of four codons.

12. The DNA construct according to claim 11 wherein said four codons of said third DNA segment are inserted at nucleotide 1152 of the gene encoding said endonuclease.

13. The DNA construct according to claim 10 wherein said third DNA segment consists essentially of 7 codons.

14. The DNA construct according to claim 13 wherein said 7 codons of said third DNA segment are inserted at nucleotide 1152 of the gene encoding said endonuclease.

15. The DNA construct according to claim 10 wherein said recognition domain is selected from the group consisting of: zinc finger motifs, homeo domain motifs, DNA binding domains of repressors, DNA binding domains of oncogenes and naturally occurring sequence-specific DNA binding proteins that recognize >6 base pairs.

16. A procaryotic cell comprising:  
(i) a first DNA segment encoding the catalytic domain of a Type IIS endonuclease which

contains the cleavage activity of said Type IIS endonuclease;

(ii) a second DNA segment encoding a sequence-specific recognition domain other than the recognition domain of said Type IIS endonuclease;

(iii) a third DNA segment comprising one or more codons, wherein said third DNA segment is inserted between said first DNA segment and said second DNA segment; and

(iv) a vector

wherein said first DNA segment, said second DNA segment, and said third DNA segment are operably linked to said vector so that a single protein is produced.

17. The procaryotic cell of claim 16 wherein said third DNA segment consists essentially of four codons.

18. The procaryotic cell of claim 16 wherein said third DNA segment consists essentially of seven codons.

19. An isolated hybrid Type IIS endonuclease produced by the procaryotic cell of claim 16.

20. An isolated DNA segment encoding the N-terminus of a Type IIS endonuclease which contains the sequence-specific recognition activity of said Type II endonuclease, said Type II endonuclease being *FokI* restriction endonuclease and having a

molecular weight of about 41 kilodaltons as measured by SDS-polyacrylamide gel electrophoresis.

21. An isolated DNA segment encoding the C-terminus of a Type IIS endonuclease which contains the cleavage activity of said Type IIS endonuclease, said Type II endonuclease being *FokI* restriction endonuclease and having a molecular weight of about 25 kilodaltons as determined by SDS-polyacrylamide gel electrophoresis.

22. An isolated protein consisting essentially of the N-terminus of the *Fok* restriction endonuclease which protein has the sequence-specific recognition activity of said endonuclease and which protein is amino acids 1-382 of said *Fok* restriction endonuclease.

23. An isolated protein consisting essentially of the C-terminus of the *FokI* restriction endonuclease which protein has the nuclease activity of said endonuclease and which protein is amino acids 383-578 of said *FokI* restriction endonuclease.

1/21

## Figure 1

FokIM5' primer

5' TA CCATGG AGGT TTAAAAT ATG AGA TTT ATT GGC AGC  
                     RBS                                    Met Arg Phe Ile Gly Ser

3' primer

18-bp complement                    NcoI  
 3' ACT ACG ACA CAG TAA ATT AAG GGTACC ATA 5'

FokIR5' primer

5' TA GGATCC GGAGGT TTAAAAT ATG GTT TCT AAA ATA AGA ACT  
                     BamHI            RBS                    Met Val Ser Lys Ile Arg Thr

3' primer

Complementary Strand                    BamHI  
 3' TTA TTG CCG CTC TAT TTG AAA ATT ACT CC TAGG AT 5'  
       Asn Asn Gly Glu Ile Asn Phe

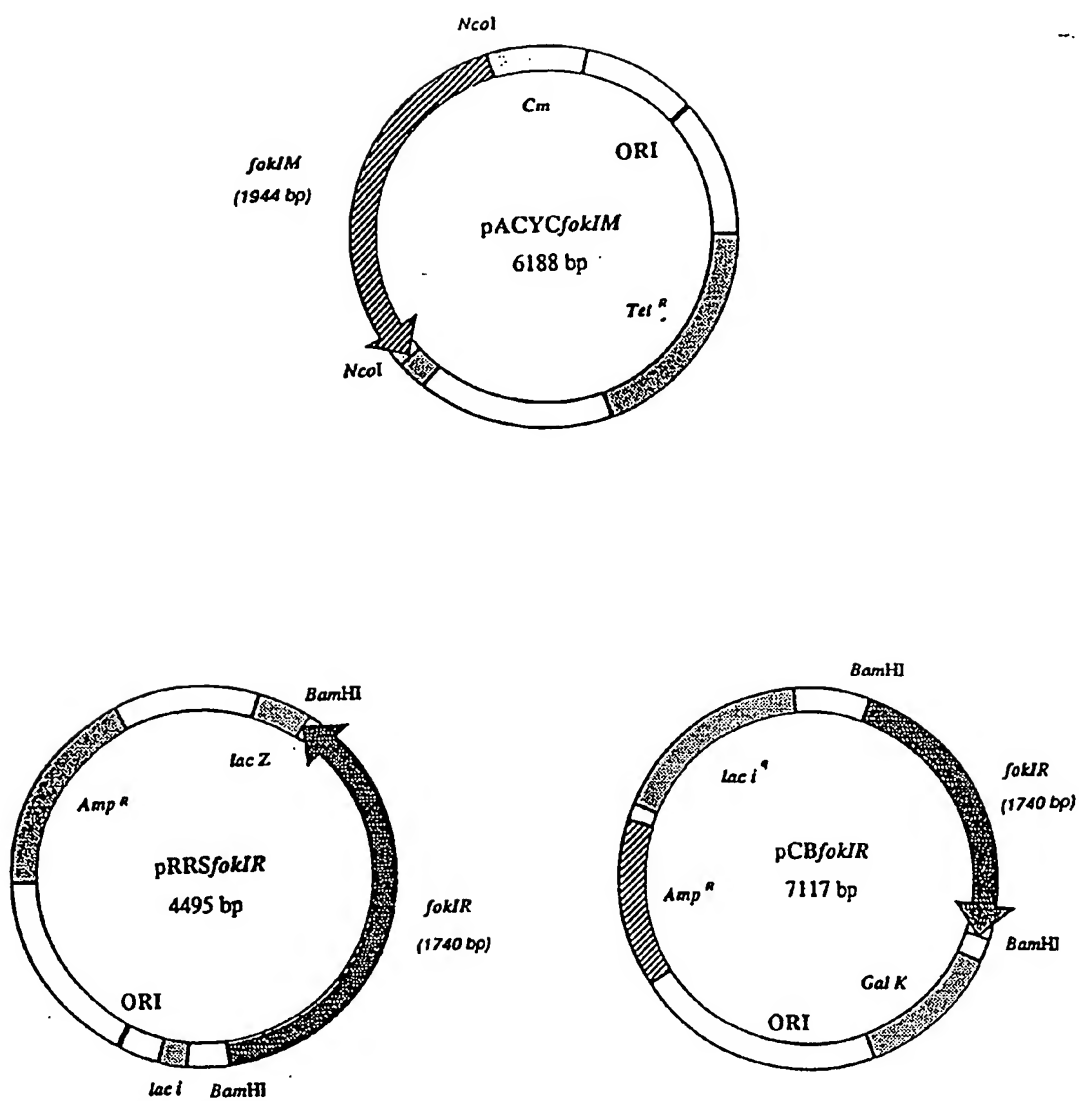
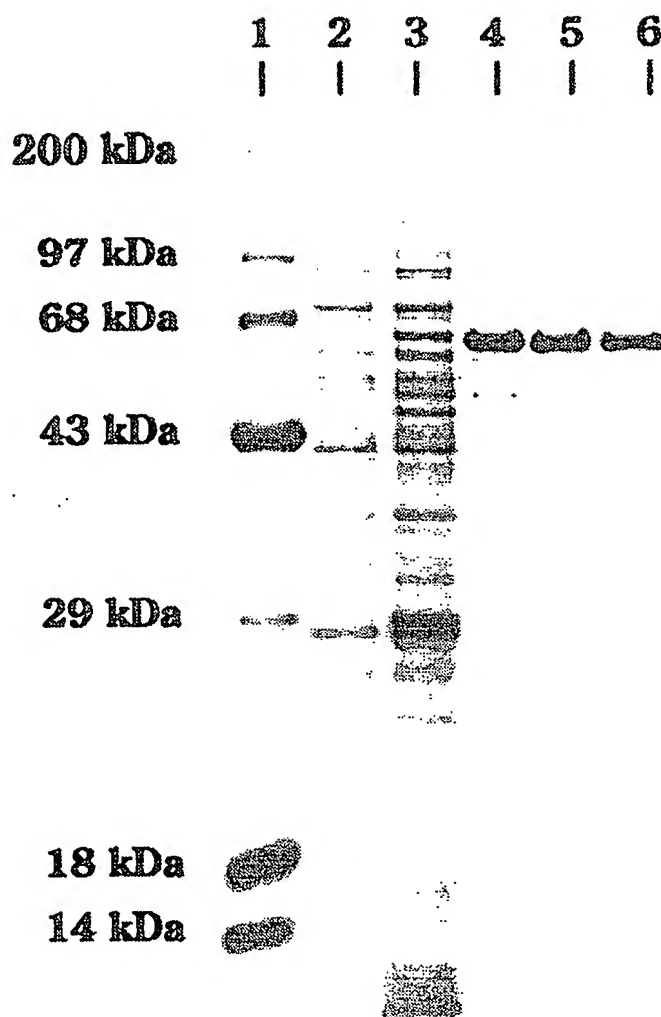


FIGURE 2

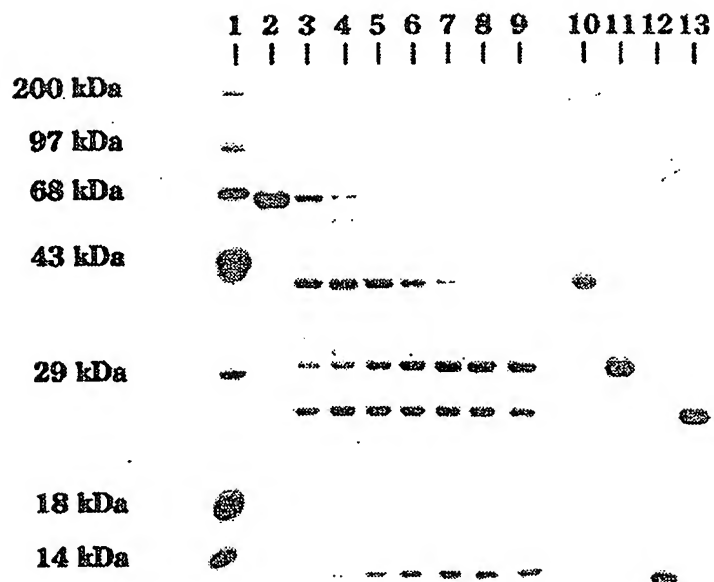
3/21

FIG. 3



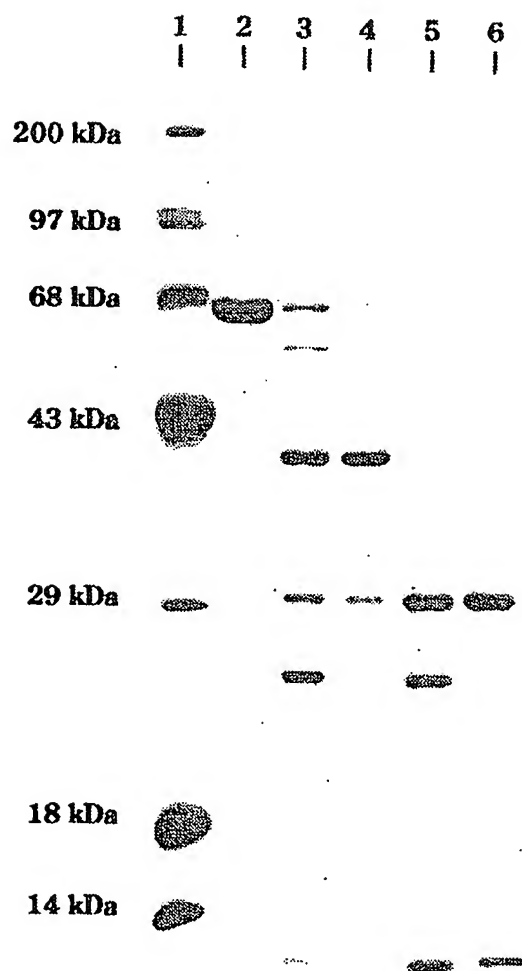
4/21

FIG. 4



5/21

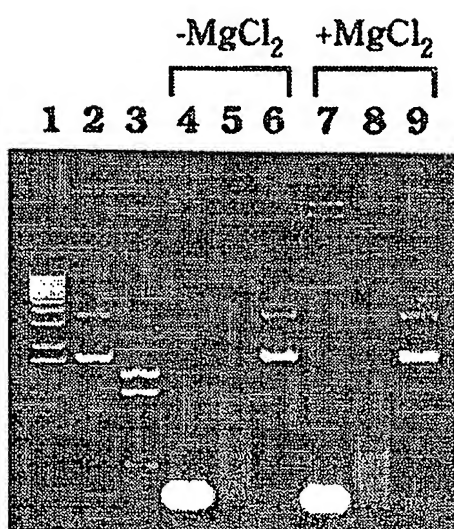
FIG. 5





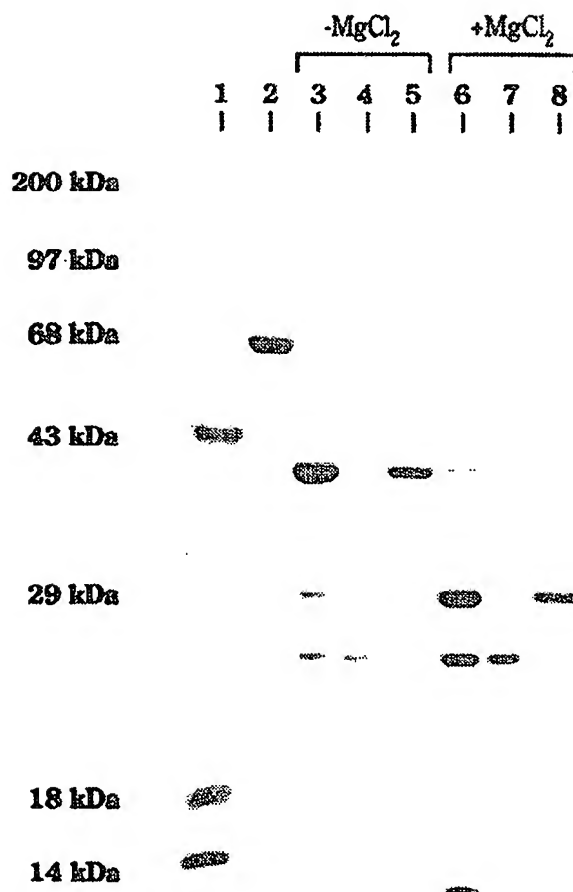
6/21

FIG. 6A



7/21

FIG. 6B



8/21

FIG.7A

FIG.7B

1 2 3 4  
| | | |

1 2 3 4  
| | | |

complex



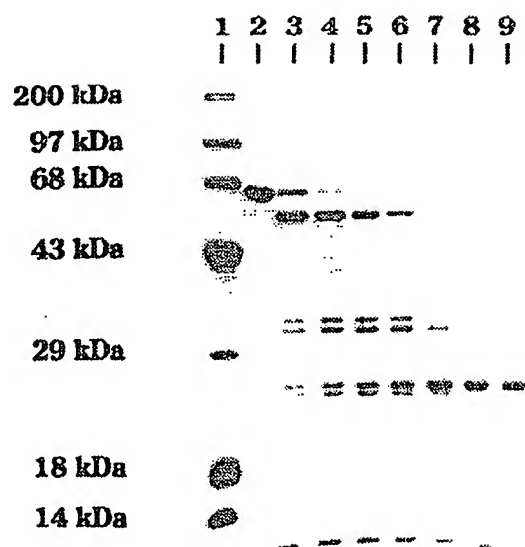
:

oligos



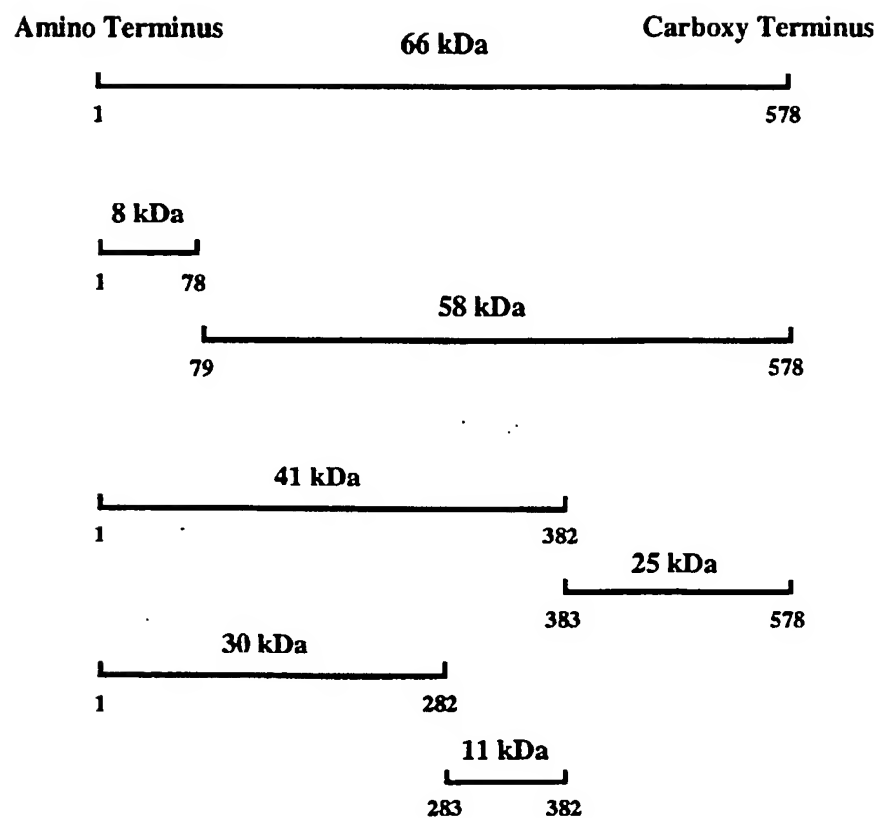
9/21

FIG. 8



10/21

FIGURE 9

***FokI* endonuclease**

11/21

**FIGURE 10**

MVSKIRTFGWQNPGEENLKRVVQVFDNRNPKVHNEVKNIKIPTLVKESKIQKELVAIMNQHDLIYTYKELVGTGTSIR  
 ss.ss.ssss.....hhhhhhhhhh.....sssssss...hhhhhhhhhhhhhhhh..ssssssssssssssssss

SEAPCDAI IQATIADQGNKKGYIDNWSSDGFRLWAHALGFIEYINKSDSFVITDVGLAYSKSADGSAIEKEILEIAISS  
 .... hhhhhhhh..... hhhhhhhh..... ssssssss..... hhhhhhhh.....

YPPAIRILTLEDGQHLTKFDLGKNLGFCGSGFTSLPEGILLDTLANAMPKDGEIRNNWEGSSDKYARMIGGWLDKL  
 .... sssshhhh..... sss..... hhhhhhhhhhh..... hhhh..... hhhh..... hhhh

GLVKQKKEFIPTLGKPDNKEFISHAFKITGEGKVLRRAGSKFTRVPKRVYWEMLATNLTDKYVTRRALILEI  
hhhhh...sss.....hhhhssss...hhhhhhhh.....hhhhhhhhhh..hhhhssshhhhhhh

LKAGSLKIEQIQDNLKKLGFEVETIENDIKGLINTGIFIEIKGRFYQLKDHILOFVIPNRGVTKQLVKSELEEKKS  
hhhh . hhhh hhhh hhhh hhhh hhhh hhhh . . ssssssssssssssssssss . . . . . hhhh hhhh hhhh hhhh

ELRHLKLYVPHEYIELIEARNSTQDRILEMKVMEFFMKVGYRGKHLGGRKPDGAITYVGSPIDYGVIVDTKAYS GG  
hhhhh . . . . hhhhhhhhhh . hhhhhhhhhh hhhhhhhh . . . . ssss . . . . sssss . . . . .

..... hhhhhhhhhh..... hhhss..... sssssss..... hhhhhhhhhhhh..... hhhhhhhh  
 YNLPIGQADEMORYVEENQTRNKHINPNEWKVPSSVTEFKLFVSGHFKGNKQAQLTRLNHITNCNGAVLSVEELLI

**GGEMIKAGTLEEVRRKFNNGEINF**  
hhhhh.hhhhhh.....

12/21

**Figure 11**

*fokIR* nt sequence      K Q L V K S E L E E K  
5'.....AAG CAA CTA GTC AAA AGT GAA CTG GAG GAG AAG ..... -3'  
Spel

5' primers:

### oligonucleotide for 4-codon insertion

5'-GGA CTA GTC AAA TCT GAA CTT AAA AGT GAA CTG GAG GAG AAG -3'

---

L V K S E L K S E L E E K

---

*Spel* 21-bp complement

### oligonucleotide for 7-codon insertion

L V K S E L E E K S E L E E K  
 5'-GGA CTA GTC AAA TCT GAA CTT GAG GAG AAG AAA AGT GAA CTG GAG GAG AAG 3'-  
 SpeI 21-bp complement

3' primer:

N F Ter Ter BamHI  
3'- TTG AAA ATT ACT CCTAGGGGCCCCCT -5'  
XmaI

FIG. 12

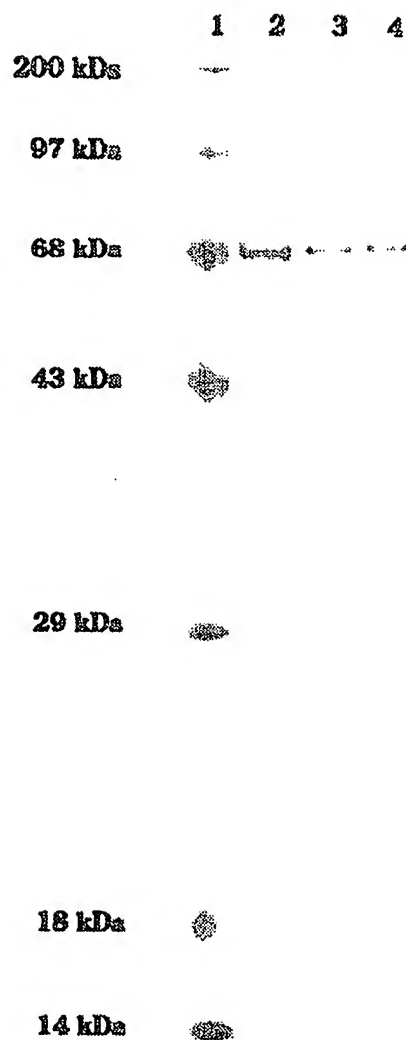




FIG. 13A

1 2 3 4 5

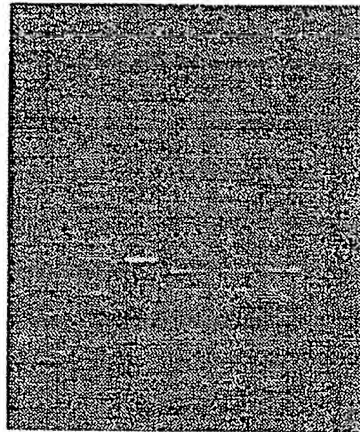


FIG. 13B

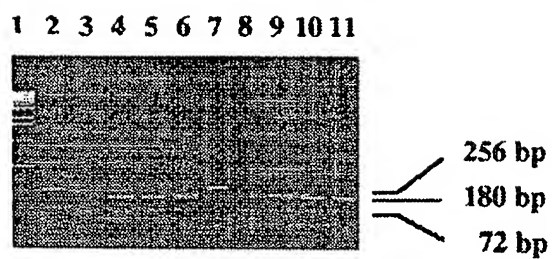


FIG. 13C

2 3 4 5 6 7 8 9 10 11



FIG. 14A

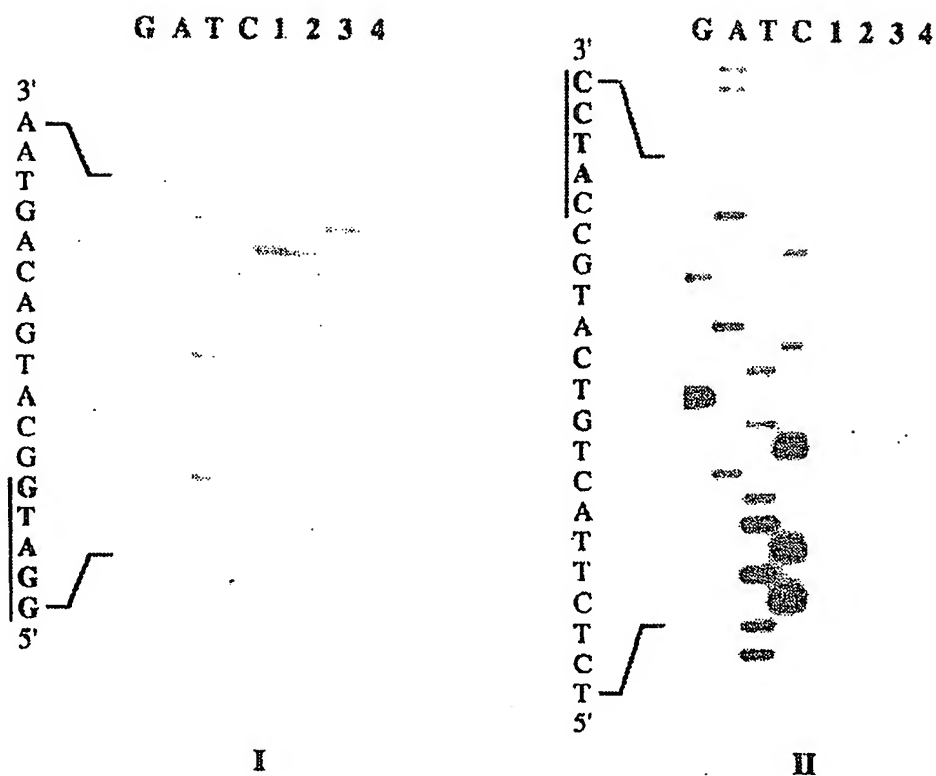
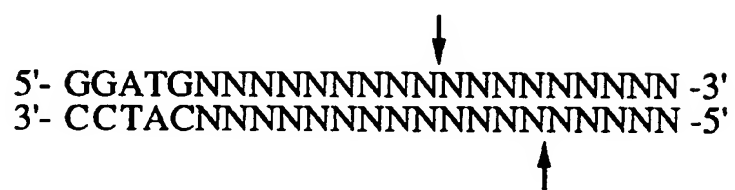


FIG. 14B



## FIG. 15A

(A) wild-type *FokI*



20/21

## FIG. 15B

**(B) 4-codon insertion mutant**

21/21

## FIG. 15C

(C) 7-codon insertion mutant





## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US94/01201

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(5) : C12N 9/22, 15/55, 15/74

US CL : 435/199, 252.3; 536/23.2

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/199, 252.3; 536/23.2; 935/47

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Computer Search - CA and APS

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	Proc. Natl. Acad. Sci, USA, Volume 89, issued May 1992, L. Li, et. al., "Functional Domains In FokI Restriction Endonuclease", pages 4275-4279, see entire document.	1-2, 4-5, 7-8, 26-27, 3, 6, 9-25, 28-29
A	Nucl. Acids Res., Volume 20, No. 16, issued 25 August 1992, K. Kita, et. al., "Cloning And Sequence Analysis Of The SstI Restriction-Modification Gene: Presence Of Homology To FokI Restriction-Modification Enzymes", pages 4167-4172, especially page 4167, column 2.	1-29



Further documents are listed in the continuation of Box C.



See patent family annex.

•	Special categories of cited documents:	T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A*	document defining the general state of the art which is not considered to be part of particular relevance		
*E*	earlier documents published on or after the international filing date	*X*	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*L*	documents which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Y*	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*O*	document referring to an oral disclosure, use, exhibition or other means		
*P*	document published prior to the international filing date but later than the priority date claimed	*A*	document member of the same patent family

Date of the actual completion of the international search

28 MARCH 1994

Date of mailing of the international search report

APR 19 1994

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

Charles Patterson

Telephone No. (703) 308-0196

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US94/01201

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Nucl. Acids Res., Volume 19, No. 5, issued 11 March 1991, H. Bocklage, et. al., "Cloning And Characterization Of The MboII Restriction-Modification System", pages 1007-1013, see especially page 1007.	1-29
A	J. Biol. Chem., Volume 264, No. 10, issued 05 April 1989, K. Kita, et. al., "The FokI Restriction-Modification System", pages 5751-5756.	1-29
A	Gene, Volume 80, issued 1989, M. C. Looney, et. al., "Nucleotide Sequence Of The FokI Restriction-Modification System; Separate Strand-Specificity Domains In The Methyltransferase", pages 193-208.	1-29